

UNIVERSIDADE FEDERAL DO PARANÁ

WILIAN CORRÊA DE MACEDO

SEQUENCIAMENTO DE ALTA COBERTURA DE REGIÕES EXONICAS DOS
GENES *PARK2*, *PACRG* E *QKI* PARA IDENTIFICAÇÃO DE VARIANTES
GENÉTICAS DE SUSCEPTIBILIDADE À DOENÇA DE PARKINSON DE INÍCIO
PRECOCE

CURITIBA
2013

WILIAN CORRÊA DE MACEDO

SEQUENCIAMENTO DE ALTA COBERTURA DE REGIÕES EXONICAS DOS
GENES *PARK2*, *PACRG* E *QKI* PARA IDENTIFICAÇÃO DE VARIANTES
GENÉTICAS DE SUSCEPTIBILIDADE À DOENÇA DE PARKINSON DE
INÍCIO PRECOCE

Dissertação apresentada ao Programa de Pós-Graduação em Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, requisito à obtenção do título de Mestre em Ciências, área de concentração Genética.

Orientador: Prof. Dr. Marcelo Távora Mira

Co-orientador: Prof. Dr. Christian Macagnan Probst

CURITIBA
2013

Universidade Federal do Paraná
Sistema de Bibliotecas

Macedo, Wilian Correa de

Sequenciamento de alta cobertura de regiões exônicas dos genes *PARK2*, *PACRG* e *QKI* para identificação de variantes genéticas de susceptibilidade à doença de parkinson de início precoce. / Wilian Correa de Macedo. – Curitiba, 2013.

137 f.; il. color. ; 30cm.

Orientador: Marcelo Távora Mira

Co-orientador: Christian Macagnan Probst

Dissertação (mestrado) - Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Parkinson, Doença de 2. Sequenciamento de próxima geração I.
Título II. Mira, Marcelo Távora III. Probst, Christian Macagnan IV.
Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa
de Pós-Graduação em Genética

CDD (20. ed.) 616.833

PARECER

Os abaixo-assinados, membros da Banca Examinadora da Defesa de Dissertação de Mestrado, a qual se submeteu **WILIAN CORRÊA DE MACEDO**, para obtenção do título de Mestre em Genética pela Universidade Federal do Paraná, no Programa de Pós-Graduação em Genética, são de parecer que se confira ao candidato o conceito "A".

Secretaria da Coordenação do Programa de Pós-Graduação em Genética do Setor de Ciências Biológicas da Universidade Federal do Paraná.

Curitiba, 8 de outubro de 2013



Doutor Christian Macagnan Probst
Fundação Oswaldo Cruz/ICC – Membro Titular



Doutor Helio Afonso Ghizoni Teive
HC/UFPR – Membro Titular

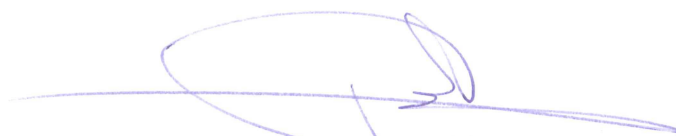


Doutora Maria Luiza Petzl-Erler
Dep. Genética/UFPR – Membro Titular



Doutor Marcelo Távora Mira
PUC/PR – Orientadora e Presidente da Banca

Visto



Professor Doutor Ricardo Lehtonen Rodrigues de Souza
Coordenador do Programa de Pós-Graduação em Genética

AGRADECIMENTOS

Agradeço ao meu orientador Prof. Dr. Marcelo T. Mira pelo acolhimento, ensinamentos, incentivo, suporte e modelo de caráter e rigor científico a ser seguido.

Ao meu co-orientador Prof. Dr. Christian M. Probst, não só por abrir as portas necessárias para a conclusão deste estudo, como também pelo suporte nas análises realizadas neste trabalho.

Aos colaboradores Prof. Dr. Hélio G. A. Teive e Prof. Dr. Renato P. Munhoz pelo profissionalismo e apoio na caracterização clínica e recrutamento da amostra populacional.

Ao Prof. Dr. Ricardo L. R. de Souza e a doutoranda Daiane P. Simão pelo suporte na expansão do grupo controle.

Aos voluntários que participaram deste trabalho por depositarem sua boa vontade e confiança em nossas mãos e por abrirem as portas de suas casas e intimidade a um, até então, completo desconhecido.

Aos meus colegas de laboratório que contribuíram com minha formação pessoal e científica durante os anos de convívio.

Aos meus pais pelo Amor incondicional e ensinamentos de vida que fizeram de mim um indivíduo consciente e humano.

Por fim, agradeço e dedico este trabalho especialmente à minha esposa Mônica Elizabeth Dallmann Sauer pela forma como seu companheirismo, força e Amor incondicional me revigoraram e me permitiram seguir em frente mesmo em tempos de dificuldade. Foi graças a ela e aos nossos incontáveis debates que mantive vivas duas características que fazem de cada pesquisador um cientista: a curiosidade e o ímpeto pela busca do conhecimento.

" No great discovery was ever made without a bold guess."

Isaac Newton

RESUMO

A Doença de Parkinson (DP) é uma desordem neurodegenerativa progressiva, de manifestação insidiosa na fase adulta tardia, causada pela morte dos neurônios dopaminérgicos da *substantia nigra pars compacta*. Quando a doença se manifesta antes dos 45 anos de idade, é caracterizada como Doença de Parkinson de Início Precoce (DPIP), uma forma de DP sob forte controle genético. Até 50% dos casos da forma autossômica recessiva de DPIP são causadas por mutações do gene *PARK2*. Inesperadamente, camundongos nocaute para *Park2* não apresentam fenótipo de parkinsonismo; por outro lado, camundongos da linhagem *Quaking*^{viable}, conhecida como modelo de desmielinização, desenvolvem tremores vigorosos por volta de 10 dias após o nascimento. Estes animais são portadores de uma deleção espontânea, homozigota recessiva, dos genes *Quaking* (*Qki*), *Park2* e *Pacrg*, vizinhos entre si e localizados no cromossomo 17A, em região homóloga à do cromossomo humano 6q25-q27. Nossa hipótese é de que mutações nos genes *QKI* e/ou *PACRG*, em conjunto ou não com mutações de *PARK2*, podem ser necessárias ou suficientes para o desenvolvimento de DPIP em humanos. Para investigar essa hipótese, análise de sequenciamento comparativo direto foi realizada com as sequências codificantes dos genes *QKI*, *PACRG* e *PARK2* de dois *pools* de DNA, um formado por trinta e seis indivíduos portadores de DPIP, e outro por dezessete indivíduos com mais de 65 anos, livres de qualquer doença neurodegenerativa, em busca de variações genéticas potencialmente patogênicas. Todas as sequências foram produzidas por sequenciamento de alta cobertura de próxima geração, conforme implementado na plataforma *Ion PGM*. Os dados primários foram submetidos a protocolo compreensivo de filtragem de artefatos de sequenciamento; as variantes que passaram pelos filtros de qualidade foram então submetidas a análises de previsão de alteração de sítios regulatórios de RNAm, de sítios de *splicing* e de sequência polipeptídica. De um total de 82 variantes identificadas, 38 foram validadas através da aplicação de critérios de qualidade (sete novas e 31 já descritas). Destas, seis variantes apresentaram evidência de consequência funcional. Duas destas variantes são intragênicas ao *PACRG*; uma *missense* com previsão de baixo impacto sobre a função do peptídeo e outra com impacto previsto sobre a regulação da expressão de *PARK2*. As quatro variantes restantes foram identificadas no gene *PARK2*. Duas delas, rs75860381 e rs1801334, já conhecidas, estão associadas com diminuição e aumento da atividade de Parquina, respectivamente – curiosamente, a primeira foi detectada apenas no *pool* pacientes, e a segunda em ambos os *pools*, mas com frequência quatro vezes maior nas *reads* do *pool* controles. Finalmente, duas variantes novas de *PARK2* foram encontradas apenas no *pool* pacientes. A primeira, uma troca C/T na posição 6:161781207, é uma mutação *nonsense* que leva à perda do sítio catalítico da parquina; a segunda é uma deleção GA/G na posição 6:161781129-6:161781128 que desencadeia um *frameshift* na área que codifica o sítio catalítico da Parquina. Em ambos os casos, a previsão é de completa abolição da atividade da enzima. Estes resultados, ainda que precisem ser validados, concordam com o corpo de evidência que sugere que variantes de *PARK2* são as únicas responsáveis pelo papel do *locus* 6q25-q27 na patogênese da DPIP.

Palavras chaves: Doença de Parkinson de início precoce. Sequenciamento de próxima geração. *PARK2*. *PACRG*. *QKI*.

ABSTRACT

Parkinson's disease (PD) is a progressive neurodegenerative disorder of insidious manifestation in late adulthood, triggered by the death of dopaminergic cells in the *substantia nigra pars compacta*. When the onset of PD occurs before 45 years of age, the disease is classified as Early Onset Parkinson's Disease (EOPD), a form of PD under strong genetic control. The *PARK2* gene has been described as causative of 50% of cases of autosomal recessive EOPD. Interestingly, an attempt to reproduce PD in *Park2* knockout mouse strain displaying no phenotype of parkinsonism. On the other hand, the *Quaking*^{viable} mice, a well known model of demyelination, develops vigorous shaking around 10 days after birth. These animals are carriers of a spontaneous, recessive deletion of *Quaking* (*Qki*), *Pacrg* and *Park2* genes, neighbors to each other and located at chromosome 17A. In humans, these genes have been mapped to chromosomal region 6q25-q27, homologous to mice chromosome 17A locus. Thus, we hypothesized that mutations in human *QKI* and/or *PACRG*, together or not with *PARK2* mutations, are necessary or sufficient for the development of EOPD. To investigate this hypothesis, direct comparative sequence analysis was performed on the coding sequences of *QKI*, *PACRG* and *PARK2*, seeking for potentially pathogenic genetic variants. The samples were composed by two DNA pools: one including 36 individuals with EOPD, and the other, 17 individuals with age higher than 65 years, free of any neurodegenerative disease. All sequences were produced by deep sequencing, as implemented in the Ion PGM platform. Primary data was subjected to a comprehensive filtering protocol designed to eliminate sequencing artifacts; variants that were approved by the quality filters were then subjected to *in silico* prediction analysis, searching for changes in regulatory sites of mRNA, splicing sites and polypeptide sequence. A total of 82 variants were identified and 38 were validated by the quality criteria (seven new and 31 known). Six out of the 38 variants showed evidence of functional consequence. Two variants are intragenic to *PACRG*; a missense variant with predicted low impact on the function of the peptide, and a variant with expected impact on the regulation of *PARK2* expression. The last four variants were identified in *PARK2*; two of them, rs75860381 and rs1801334, have been described in the public databases and have been associated with a decreased and increased parkin activity, respectively. Interestingly, the first variant was detected only in the patient pool, and the second in both pools, but the frequency was four times higher in the reads of the control pool. Finally, two new variants of *PARK2* were found only in the patient pool. The first, a C/T change in position 6:161781207, generates a nonsense mutation which leads to the loss of catalytic site of parkin; the second is a GA/G deletion at 6:161781129-6:161781128 which creates a frameshift in the coding area of the catalytic site of parkin. In both cases, the functional prediction is the complete abolishment of enzyme activity. These results, yet to be validated by another sequencing technique, agree with the cumulative evidence suggesting that variants of *PARK2* alone are responsible for the role of the 6q25-q27 locus in the pathogenesis of EOPD.

Keywords: Early Onset Parkinson's disease. Next-generation sequencing. *PARK2*. *PACRG*. *QKI*.

LISTA DE ILUSTRAÇÕES

Figura 1	Representação de uma região de aproximadamente 2,5Mb no cromossomo 6q humano.....	31
Figura 2	Visualização de alinhamento de <i>reads</i> e apresentação dos principais indicadores, no programa IGV.....	40
Figura 3	Desenho geral do estudo.....	46
Figura 4	Esquema de análise dos dados.....	57
Figura 5	Visão esquemática do alinhamento e cobertura de duas regiões com qualidades diferentes segundo o algoritmo IGV.....	63
Figura 6	Imagens das áreas de carregamento contendo as microesferas (<i>Ion Sphere Particles</i> - ISP) com os <i>pools</i> de amplicons.....	67
Figura 7	Rendimento das microesferas (<i>Ion Sphere Particles</i> - ISP) geradas....	68
Figura 8	Distribuição dos tamanhos das leituras geradas.....	69
Figura 9	Estatística do total de <i>reads</i> mapeadas.....	70
Figura 10	Visão esquemática do alinhamento e cobertura das regiões exônicas sequenciadas dos genes <i>PACRG</i> , <i>PARK2</i> e <i>QKI</i> segundo o algoritmo IGV.....	73
Figura 11	Análise de conservação das sequências de aminoácidos da proteína parquína entre as espécies <i>Homo sapiens</i> , <i>Pan troglodytes</i> , <i>Mus musculus</i> , <i>Rattus norvegicus</i> , <i>Gallus gallus</i> , <i>Danio rerio</i> e <i>Drosophila melanogaster</i>	86
Figura 12	Análise de conservação das sequências de aminoácidos da proteína PACRG entre as espécies <i>Homo sapiens</i> , <i>Pan troglodytes</i> , <i>Mus musculus</i> , <i>Rattus norvegicus</i> , <i>Gallus gallus</i> , <i>Danio rerio</i> e <i>Drosophila melanogaster</i>	88
Figura 13	Análise de conservação das sequências de aminoácidos da proteína QKI entre as espécies <i>Homo sapiens</i> , <i>Pan troglodytes</i> , <i>Mus musculus</i> , <i>Rattus norvegicus</i> , <i>Gallus gallus</i> , <i>Danio rerio</i>	90
Figura 14	Apresentação da estrutura primária da parquína na sua forma nativa e alterada por duas mutações.....	97

LISTA DE TABELAS

Tabela 1	Critérios para diagnóstico clínico de DP segundo o Banco de Cérebros da Sociedade de Parkinson do Reino Unido.....	17
Tabela 2	<i>Loci</i> PARK1 a 18 com suas respectivas localizações cromossômicas e principais genes associados (quando identificados).....	26
Tabela 3	Características demográficas e clínicas da amostra populacional...	66
Tabela 4	Qualidade de alinhamento das <i>reads</i>	71
Tabela 5	Número de <i>reads</i> e bases mapeados e suas respectivas coberturas (extensão) na referência.....	72
Tabela 6	Extensão da referência e dados de profundidade de cobertura.....	72
Tabela 7	Número de variantes novas e já descritas encontradas nos <i>pools</i> em estudo para <i>PACRG</i> , <i>PARK2</i> e <i>QKI</i>	75
Tabela 8	Variantes identificadas já descritas em bancos de dados que são compartilhadas entre pacientes e controles	76
Tabela 9	Variantes identificadas já descritas em bancos de dados presentes somente em controles.....	77
Tabela 10	Variantes identificadas já descritas em bancos de dados presentes somente em pacientes.....	77
Tabela 11	Variantes novas que são compartilhadas entre pacientes e controles.....	79
Tabela 12	Variantes novas presentes somente em controles.....	80
Tabela 13	Variantes novas presentes somente em pacientes.....	81
Tabela 14	Variantes já descritas e filtradas: identificação segundo o dbSNP, base de dados em que se encontram e predição de impacto funcional.....	83
Tabela 15	Variantes novas filtradas e previsão de impacto funcional.....	85

LISTA DE ABREVIATURAS E SIGLAS

μL	Microlitros
18F-DOPA	L-6-fluoro 3,4-dihydroxyphenylalanine
A	Adenina
A	Alanina
AA	Aminoácido
APPP	Associação Paranaense de Portadores de Parkinson
AQ17	<i>Average quality 17</i>
AQ20	<i>Average quality 20</i>
ATP13A2	<i>ATPase type 13A2</i>
AVE	<i>Acidente vascular encefálico</i>
AVG	<i>Average</i>
BST1	<i>Bone marrow stromal cell antigen 1</i>
C	Cisteína
C	Citosina
Chr	<i>Chromosome</i>
CNV	<i>Copy Number Variation</i>
COMT	<i>Catecol-O-methyltransferase</i>
D	Ácido aspártico
DCMVR	Doença Comum – Múltiplas Variantes Raras
Dir	Direta
dbSNP	<i>Single Nucleotide Polymorphism Database</i>
ddH ₂ O	<i>Água bidestilada</i>
DNA	<i>Deoxyribonucleic acid</i>
dNTP	Desoxinucleotídeo trifosfatado
DP	Doença de Parkinson
DPIP	Doença de Parkinson de Início Precoce
DPIP-AR	Doença de Parkinson de Início Precoce Autossômico Recessivo
DPIT	Doença de Parkinson de Início Tardio
Dr	Doutor
DZ	Dizigótico
E	Ácido glutâmico
EDTA	Ethylenediaminetetraacetic acid
EIF4G1	<i>Eukaryotic translation initiation factor 4 gamma, 1</i>
EOPD	<i>Early Onset Parkinson's Disease</i>
ePCR	<i>Emulsion PCR</i>
Et al	<i>Et alii</i>
F	Fenilalanina
FBXO7	<i>F-box protein 7</i>
G	Glicina
G	Guanina
GABA	<i>Gamma-AminoButyric Acid</i>

GAK	<i>Cyclin G associated kinase</i>
GIGYF2	<i>GRB10 interacting GYF protein 2</i>
GRCh37.p10	<i>Genome Reference Consortium human build 37, patch release 10</i>
GWAS	<i>Genome-Wide Association Study</i>
H	Histidina
HapMap	<i>Haplotype map</i>
HC	Hospital de Clínicas da UFPR
HGMD	<i>Human Gene Mutation Database</i>
HLA-DRB5	<i>Major histocompatibility complex, class II, DR beta 5</i>
HTRA2	<i>HtrA serine peptidase 2</i>
Hz	Hertz
I	Isoleucina
IBR	<i>In-Between-RING</i>
IC 95%	Intervalo de confiança de 95%
ICC	Instituto Carlos Chagas
IGV	<i>Integrative Genome Viewer</i>
IL1B	<i>Interleukin 1-beta</i>
INDEL	Inserção/deleção
IPDGC	<i>International Parkinson Disease Genomics Consortium</i>
ISFET	<i>Ion-sensitive field-effect transistor</i>
ISP	<i>Ion Sphere Particles</i>
K	Lisina
Kb	Kilobases
KH domain	<i>K homology domain</i>
L	Leucina
L	Litro
LOD	<i>Logarithm (base 10) of odds</i>
LOVD	<i>Leiden Open Variation Database</i>
LRRK2	<i>Leucine-rich repeat kinase 2</i>
M	Metionina
M	Molar
MAOB	<i>Monoamine oxidase B</i>
MAPT	<i>Microtubule-associated protein tau</i>
Mb	Mega base
mer	Unidade de polímero
miRNA	MicroRNA
mL	Mililitros
MPTP	1-metil-4-fenil-1,2,3,6-tetrahidropiridina
mRNA	<i>Messenger RNA</i>
MS	Ministério da Saúde
MUSCLE	<i>Multiple Sequence Comparison by Log-Expectation</i>
MZ	Monizigótico
N	Aspargina
N	Número
NAT2	<i>N-acetyltransferase 2</i>
NCBI	<i>National Center for Biotechnology Information</i>

ng	Nanograma
NGRC	<i>NeuroGenetics Research Consortium</i>
NGS	<i>Next-generation sequencing</i>
nm	Nanômetro
nM	Nanomolar
NOS	<i>Nitric oxide synthase</i>
<i>NURR1</i>	<i>Nuclear receptor related 1</i>
OMIM	<i>Online Mendelian Inheritance in Man</i>
OR	<i>Odds Ratio</i>
P	Prolina
<i>PACRG</i>	<i>PARK2 co-regulated</i>
<i>PARK2</i>	<i>Parkinson protein 2</i>
pb	Pares de bases
PCR	<i>Polymerase Chain Reaction</i>
PD	<i>Parkinson's Disease</i>
PET	<i>Positron emission tomography</i>
Pfam	<i>Protein families</i>
PGM	<i>Personal Genome Machine</i>
<i>PINK1</i>	<i>PTEN induced putative kinase 1</i>
<i>PLA2G6</i>	<i>Phospholipase A2, group VI</i>
pM	Picomolar
pmol	Picomol
PolyPhen-2	<i>Polymorphism phenotyping 2</i>
Prof	Professor
PUCPR	Pontifícia Universidade Católica do Paraná
Q	Glutamina
<i>QKI</i>	<i>Quaking</i>
<i>Qk^v</i>	<i>Quaking^{viable}</i>
qsp	Quantidade suficiente para
Q_Var	Qualidade de identificação de Variante
R	Arginina
Ref	Referência
REP	<i>Repressor Element of Parkin</i>
Rev	Reversa
RING	<i>Really Interesting New Gene</i>
RNA	<i>Ribonucleic acid</i>
RNA _m	<i>RNA mensageiro</i>
RR	Risco relativo
S	Serina
SAS	Secretaria de Atenção à Saúde
SDS	<i>Sodium dodecyl sulfate</i>
SIFT	<i>Sorting intolerant from tolerant</i>
SNC	Sistema Nervoso Central
SNCA	<i>Synuclein alpha</i>
SNCAIP	<i>Synuclein, alpha interacting protein</i>

SNP	<i>Single nucleotide polymorphism</i>
SNPC	<i>Substantia nigra pars compacta</i>
SNV	<i>Single nucleotide variant</i>
SOLiD	<i>Sequencing by Oligonucleotide Ligation and Detection</i>
T	Timina
T	Treonina
TCLE	Termo de consentimento livre e esclarecido
TE	Solução de Tris e EDTA
<i>TFAM</i>	<i>Transcription factor A, mitochondrial</i>
TKM1	Solução de Tris, cloreto de potássio e cloreto de magnésio
TKM2	Solução TKM1 acrescida de cloreto de sódio
T _m	<i>Melting temperature</i>
TMAP	Torrent Mapping Alignment Program
TVC	Torrent variant caller
U	Unidade
UBL	<i>Ubiquitin-like</i>
<i>UCHL1</i>	<i>Ubiquitin carboxyl-terminal esterase L1</i>
UCSC	<i>University of California, Santa Cruz</i>
UFPR	Universidade Federal do Paraná
UTR	<i>Untranslated region</i>
V	Valina
Var	Variante
<i>VPS35</i>	<i>Vacuolar protein sorting 35 homolog</i>
W	Triptofano
X	Vezes
Y	Tirosina

SUMÁRIO

1	INTRODUÇÃO	15
1.1	DOENÇA DE PARKINSON – ASPECTOS GERAIS	15
1.1.1	Características clínicas e diagnóstico	16
1.1.2	Fisiopatologia e patogênese	17
1.1.3	Tratamento	19
1.2	GENÉTICA DA DOENÇA DE PARKINSON	20
1.2.1	Estudos em gêmeos e herdabilidade da DP	20
1.2.2	Modelos de herança para DP	23
1.2.3	Estudos de ligação, associação e sequenciamento	24
1.3	UMA NOVA PERSPECTIVA SOBRE A REGIÃO 6q25-q27: <i>PARK2</i> É O ÚNICO RESPONSÁVEL PELA LIGAÇÃO DO <i>LOCUS</i> <i>PARK2</i> À DOENÇA DE PARKINSON DE INÍCIO PRECOCE?	27
1.4	DESCRIÇÃO DA ESTRUTURA E FUNÇÕES DOS GENES EM ESTUDO	33
1.4.1	<i>PARK2</i>	33
1.4.2	<i>PACRG</i>	34
1.4.3	<i>QKI</i>	35
1.5	APLICAÇÃO DE PLATAFORMA DE SEQUENCIAMENTO DE PRÓXIMA GERAÇÃO PARA A IDENTIFICAÇÃO DE VARIANTES CAUSAIS	35
1.5.1	Bases conceituais: plataforma <i>Ion Torrent PGM</i> e características de dados NGS	37
1.5.1.1	Biblioteca de DNA	37
1.5.1.2	PCR em emulsão (ePCR)	37
1.5.1.3	Princípio do sequenciamento pela plataforma <i>Ion Torrent PGM</i>	38
1.5.1.4	Características dos dados NGS	39
1.5.2	Extratécnica de <i>pooling</i> para detecção de variantes	41
2	JUSTIFICATIVA	43
3	OBJETIVOS	44
4	MATERIAIS E MÉTODOS	45
4.1	INSTITUIÇÃO DE ORIGEM E INSTITUIÇÕES PARCEIRAS	45
4.2	DESENHO DE ESTUDO	45
4.3	RECRUTAMENTO E CARACTERIZAÇÃO CLÍNICA DA AMOSTRA POPULACIONAL	47
4.4	COLETA DE AMOSTRA E EXTRAÇÃO DE DNA	47
4.5	QUANTIFICAÇÃO E PADRONIZAÇÃO DE AMOSTRAS	48
4.6	DESENHO DOS INICIADORES (<i>PRIMERS</i>)	48
4.7	AMPLIFICAÇÃO DE DNA	49
4.8	QUANTIFICAÇÃO, PADRONIZAÇÃO DE SOLUÇÃO DE AMPLICONS E <i>POOLING</i>	49

4.9	CONCATENAÇÃO DE AMPLICONS	50
4.10	PREPARO DAS BIBLIOTECAS	50
4.10.1	Fragmentação do DNA	51
4.10.2	Ligação dos adaptadores e <i>nick-repair</i>	52
4.10.3	Seleção de tamanho dos fragmentos.....	52
4.10.4	Amplificação e purificação das bibliotecas	53
4.10.5	Quantificação e diluição das bibliotecas amplificadas	53
4.11	PREPARO DOS MOLDES	53
4.11.1	PCR em emulsão (ePCR).....	54
4.11.2	Quebra da emulsão e lavagem das microesferas	55
4.11.3	Enriquecimento das microesferas carregadas	55
4.12	SEQUENCIAMENTO	56
4.13	ANÁLISE DOS DADOS.....	56
4.13.1	Análise primária	58
4.13.2	Análise secundária	58
4.13.3	Análise terciária	64
5	RESULTADOS E DISCUSSÃO	66
5.1	DESCRIÇÃO DA POPULAÇÃO.....	66
5.2	ANÁLISE DOS DADOS.....	66
5.2.1	Análise primária	66
5.2.2	Análise secundária	69
5.2.3	Análise terciária	81
5.2.3.1	Análise de impacto funcional	81
5.2.3.2	Análise de conservação	85
5.2.3.3	Interpretação dos resultados	91
5.3	LIMITAÇÕES	97
5.3.1	Limitações do estudo	97
6	CONCLUSÕES	99
	REFERÊNCIAS.....	101
	APÊNDICE A - TERMO DE CONCENTIMENTO LIVRE E ESCLARECIDO.....	125
	APÊNDICE B - FERRAMENTAS ONLINE UTILIZADAS	129
	APÊNDICE C - <i>PRIMERS</i> DESENHADOS PARA OS GENES <i>PACRG</i> , <i>PARK2</i> E <i>QKI</i>	130
	APÊNDICE D - TEMPERATURA EXPERIMENTAL DE HIBRIDAÇÃO DOS PARES DE <i>PRIMERS</i>	133
	ANEXOS	134
	ANEXO A - EXTRAÇÃO DE DNA PELA TÉCNICA DE <i>SALTING OUT</i>	134

1 INTRODUÇÃO

1.1 DOENÇA DE PARKINSON – ASPECTOS GERAIS

A Doença de Parkinson (DP; OMIM #168600), descrita pela primeira vez em 1817 por James Parkinson, é uma desordem neurodegenerativa progressiva e geralmente de manifestação insidiosa na fase adulta tardia. Em seu trabalho, Parkinson descreveu, em seis pacientes, a maioria das características clínicas típicas da doença, tais como bradicinesia, tremor em repouso e instabilidade postural (1). O relato de Parkinson teve pouco impacto até que, em 1888, Jean-Martin Charcot, reconhecendo a importância de seu predecessor, deu continuidade ao trabalho, adicionando a rigidez postural à lista de características clínicas da doença. Charcot propôs ainda batizar a doença com o nome do seu primeiro relator (2). Atualmente, a DP ocupa a segunda posição em prevalência entre as doenças neurodegenerativas comuns e, desta categoria, é a mais frequente das desordens do movimento, afetando por volta de 2% da população mundial com idade superior a 65 anos e mais de 4% dos idosos acima de 85 anos (3-5). Pesquisa realizada em 2006 por Barbosa *et al.* em uma amostra populacional brasileira revelou 3,3% de taxa de prevalência da DP em pacientes com idades superiores a 64 anos (6).

A DP tem incidência maior em homens do que em mulheres (3, 7) e os primeiros sintomas podem ter início entre 40 e 70 anos, sendo mais frequente aos 60 anos (2, 8). Os pacientes que apresentam manifestações clínicas da DP antes dos 45 anos tem sua desordem classificada como Doença de Parkinson de Início Precoce (DPIP), que responde de 5 a 10% do total de pacientes com DP (9, 10). Inicialmente, a idade de corte para caracterização da DPIP foi de 40 anos (11); no entanto, com a identificação de mutações de DNA em uma parcela expressiva de pacientes com DPIP cujo sintomas manifestavam-se mais tardiamente, a idade de corte foi reajustada para 45 anos (12, 13).

1.1.1 Características clínicas e diagnóstico

Na DP, as principais características clínicas continuam sendo aquelas já descritas por Parkinson e Charcot: bradicinesia, tremor em repouso, instabilidade postural e rigidez, um quadro caracterizado como síndrome parkinsoniana ou parkinsonismo (14). Além delas, os pacientes de DP desenvolvem perda dos movimentos autônomos, como balanço do braço ao caminhar, controle de piscar e deglutição, disfunção da marcha e dificuldades de fala. Ainda, sintomas não-motores são frequentes durante o curso da doença, tais como comprometimento cognitivo (em até 20% dos pacientes), alucinações, delírio, distúrbios de comportamento, depressão, distúrbios de sono e vigília, déficit olfatório e insuficiência do sistema autônomo resultando em quadros de constipação, hipotensão, impotência e sudorese (5, 15-19).

O diagnóstico da doença é essencialmente clínico e, embora heterogênea, a manifestação clínica da DP tem sido tradicionalmente definida pela presença de sinais motores (5, 16). Atualmente, os critérios do Banco de Cérebros da Sociedade de Parkinson do Reino Unido (*United Kingdom Parkinson Disease Society Brain Bank*) são os mais utilizados para o diagnóstico de DP (20). Com base neste consenso, o paciente terá diagnóstico de DP se apresentar dois critérios necessários, sendo um deles obrigatoriamente bradicinesia, e pelo menos um dos três critérios de suporte positivo (TABELA 1).

A caracterização da DPIP tem como único diferencial, em relação à forma clássica, a manifestação da síndrome parkinsoniana antes dos 45 anos de idade (10).

TABELA 1. Critérios para diagnóstico clínico de DP segundo o Banco de Cérebros da Sociedade de Parkinson do Reino Unido.

Critérios para diagnóstico de Doença de Parkinson	
Critérios necessários para diagnóstico de DP (pelo menos 2)	Critérios negativos para DP (excludentes)
<ol style="list-style-type: none"> 1. Bradicinesia (<i>obrigatório</i>), 2. Rigidez muscular, 3. Tremor de repouso (4-6 Hz) avaliado clinicamente, 4. Instabilidade postural não causada por distúrbios visuais, vestibulares, cerebelares ou proprioceptivos. 	<ol style="list-style-type: none"> 1. História de AVE de repetição, 2. História de trauma craniano grave, 3. História definida de encefalite, 4. Crises oculogíricas, 5. Tratamento prévio com neurolépticos, 6. Remissão espontânea dos sintomas, 7. Quadro clínico estritamente unilateral após 3 anos, 8. Paralisia supranuclear do olhar, 9. Sinais cerebelares, 10. Sinais autonômicos precoces, 11. Demência precoce, 12. Liberação piramidal com sinal de Babinski, 13. Presença de tumor cerebral ou hidrocefalia comunicante, 14. Resposta negativa a altas doses de levodopa, 15. Exposição a 1-metil-4-fenil-1,2,3,6-tetrahidropiridina (MPTP)
Critérios de suporte positivo para o diagnóstico de DP (pelo menos 3 são necessários para o diagnóstico)	
<ol style="list-style-type: none"> 1. Início unilateral, 2. Presença de tremor de repouso, 3. Doença progressiva, 4. Persistência da assimetria dos sintomas, 5. Boa resposta a levodopa, 6. Presença de discinesias induzidas por levodopa, 7. Resposta a levodopa por 5 anos ou mais, 8. Evolução clínica de 10 anos ou mais. 	

FONTE: Adaptado de Hughes, A.J. *et al.* 1992 (20) e Portaria SAS/MS nº 228, 2010(21).

1.1.2 Fisiopatologia e patogênese

A etiologia da doença ainda é desconhecida; contudo, supõe-se a participação de vários mecanismos que, em conjunto, são capazes de desencadeá-la. Um corpo crescente de evidências provenientes de estudos de fatores de risco ambientais e genéticos aponta para uma convergência entre metabolismo energético e processamento de proteínas danificadas no desenvolvimento da DP (22). Esses dados sugerem que alterações na função mitocondrial e do sistema ubiquitina-proteasomal contribuem para a patogênese da doença (22-24). Outros fatores que

têm sido implicados na degeneração neuronal da DP são estresse oxidativo, ação de toxinas excitatórias e deficiência de fatores neurotróficos (22).

A síndrome parkinsoniana, principal manifestação clínica da DP, é decorrente do comprometimento da via dopaminérgica nigro-estriatal. Sabe-se que a disfunção motora é resultado de uma progressiva perda de células dopaminérgicas na área *substantia nigra pars compacta* (SNPC), situada no mesencéfalo, e torna-se evidente clinicamente quando aproximadamente 70 a 80% dos neurônios dopaminérgicos são perdidos, o que corresponde a cerca de 450.000 células dopaminérgicas da SNPC em um adulto (25). Um fato curioso é que as disfunções motoras não são evidentes até que haja uma perda de 80% do aporte de dopamina no corpo estriado (núcleo caudado e putâmen), o que sugere um mecanismo compensatório muito eficiente nos estágios iniciais da doença (5, 16, 26).

Achados histopatológicos na forma clássica da doença, além da perda de neurônios dopaminérgicos, incluem a presença de corpúsculos de Lewy em neurônios sobreviventes em várias áreas do cérebro (27-30). Corpúsculos de Lewy são inclusões citoplasmáticas formadas pelo acúmulo de proteínas de diversas classes, com destaque para proteases, proteínas do citoesqueleto, grande quantidade de alfa-sinucleína e produtos dos genes *PARK2* e *PACRG*; as proteínas componentes desta estrutura podem estar conjugadas ou não com ubiquitina (29-31). No entanto, estudos *post-mortem* de casos de DPIP mostram um perfil atípico de DP em que os corpúsculos de Lewy estão ausentes. Esta ocorrência é verificada em pacientes com a forma autossômica recessiva de início precoce da doença associada com mutações de *PARK2* (16, 28, 32, 33).

Atualmente, não há consenso se os corpúsculos de Lewy são causa ou consequência da DP, com evidências sugerindo que eles desempenham um papel citoprotetor sequestrando oligômeros tóxicos de alfa-sinucleína. Esta molécula faz parte de uma família de proteínas de expressão abundante no cérebro e, provavelmente, tem a função de integrar eventos de sinalização pré-sinápticos ao tráfego de moléculas pelas membranas em neurônios (34). No entanto, evidências acumuladas de estudos *in vitro* (35) e em modelos animais (36-38) sugerem que tanto agregados de oligômero quanto excesso de formas fibrilares de alfa-sinucleína são tóxicas para os neurônios (39). Finalmente, evidências acumuladas indicam que processos celulares como estresse oxidativo, disfunção mitocondrial e remoção de

proteínas deformadas são processos envolvidos no mecanismo de patogênese da DP, contudo não se sabe quais destes mecanismos são causa ou consequência da desordem (5, 39).

1.1.3 Tratamento

Até o momento, não há cura para a DP; o tratamento farmacológico existente tem como propósito restabelecer as funções motoras e aliviar condições associadas à doença. Levodopa ainda é o tratamento quimioterápico mais efetivo, mas seu uso é complicado pelo aparecimento de instabilidade motora e discinesia. Outras opções de tratamento incluem agonistas dopaminérgicos, anticolinérgicos, amantadina, inibidores da monoamino oxidase e inibidores da catecol-O-metiltransferase (40, 41). A escolha do medicamento mais adequado leva em consideração fatores como estágio da doença, sintomatologia presente, ocorrência de efeitos colaterais, idade do paciente, medicamentos em uso e seu custo (21).

Estima-se que a taxa de morte dos neurônios dopaminérgicos na forma idiopática da doença seja de aproximadamente 10% ao ano (42). Consequentemente, com o tempo, os sintomas de parkinsonismo pioram e há necessidade de se aumentar a dose e/ou associar medicações para controle dos sintomas. O grau de resposta aos medicamentos vai decrescendo com a progressão da doença e novos sintomas vão surgindo, acarretando em grave incapacidade após 10 a 15 anos (21). Alternativa aos métodos farmacológicos tem sido a terapia de estimulação cerebral profunda (*deep brain stimulation*), que consiste em um procedimento cirúrgico de implante de um eletrodo em áreas específicas do cérebro. No caso de DP, os alvos são o núcleo subtalâmico e o globo pálido, áreas que passam a ter uma atividade excessiva e anormal frente a perda dos neurônios dopaminérgicos. Este procedimento não visa correção direta da falta de dopamina, mas sim equilibrar o excesso de atividades dos núcleos através de pulsos elétricos contínuos (43, 44).

1.2 GENÉTICA DA DOENÇA DE PARKINSON

A hipótese da existência de fatores genéticos envolvidos no controle da susceptibilidade do ser humano à DP e ao parkinsonismo vem sendo tema de intensa investigação. Embora estudos iniciais, incluindo estudos de gêmeos, indicavam que fatores ambientais eram mais importantes que fatores genéticos no controle da susceptibilidade à DP (45-50), hoje é amplamente aceita a existência de um componente genético envolvido na patogênese da doença, com os primeiros relatos da existência de pacientes com parentes afetados datados do início do século XX (51). Desde então, agregação de casos de DP em famílias tem sido reportada em vários trabalhos, sendo que aproximadamente 10% dos pacientes relataram um histórico familiar positivo; muitos destes casos tem em comum o aparecimento precoce da doença (52-54).

Embora existam formas de DP com clara etiologia genética, a maior parte dos casos são idiopáticos, ou seja, uma DP multifatorial, provavelmente resultante da combinação de herança poligênica, variáveis ambientais e interações de genes de susceptibilidade com o ambiente (5, 55, 56). Em termos de fatores de risco não-genéticos, há exemplos de degeneração neuronal na SNPC induzidos por intoxicação com 1-metil-4-fenil-1,2,3,6-tetrahidropiridina (MPTP) (46, 57) e parkinsonismo pós-encefálico (58), mas nenhum deles reproduz com fidelidade as características clínicas e patológicas da doença de Parkinson verdadeira (5). Além disso, estudos epidemiológicos demonstraram que a exposição a pesticidas (59) e metais pesados (60) estão associados à DP e que o hábito de fumar (61-64) e o consumo de café ou chá (65, 66) diminuem o risco de desenvolver a desordem.

1.2.1 Estudos em gêmeos e herdabilidade da DP

Estudo de gêmeos foi o desenho inicial de investigação, visando distinguir a

contribuição genética dos fatores de risco ambientais para a manifestação da DP (47, 67, 68). Por exemplo, em 1999, Tanner *et al.* (49) realizaram um estudo em uma amostra populacional de 71 gêmeos monozigóticos (MZ) e 90 dizigóticos (DZ) com pelo menos um irmão com diagnóstico de DP; como resultado obtiveram concordância similar quando a análise envolvia a população total (16% e 11% respectivamente - risco relativo [RR]: 1,39; intervalo de confiança de 95% [IC 95%]: 0,63-3,1). No entanto, ao analisar apenas os casos com início da doença antes dos 50 anos (4 gêmeos MZ e 12 DZ), a taxa de concordância foi significativamente maior em gêmeos MZ (100% em MZ comparados com 16% em DZ - RR: 6,0; IC 95%: 1,69-21,26). Os autores sugeriram que entre indivíduos com DP de início tardio, fatores de risco ambientais são mais importantes, enquanto que para a forma precoce, fatores genéticos parecem ter maior relevância.

Contrastando com o estudo anterior, uma pesquisa de desenho longitudinal e funcional publicada no mesmo ano obteve resultado interessante ao avaliar disfunção nigroestriatal usando 18F-DOPA e tomografia por emissão de pósitrons (do inglês *positron emission tomography*, PET) em gêmeos aparentemente discordantes para DP de início tardio (69). Eles observaram uma maior concordância entre 18 gêmeos MZ do que em 16 DZ em relação ao fenótipo de captação competitiva de 18F-DOPA (55% e 18%, respectivamente). Após 7 anos de acompanhamento, todos os irmãos monozigóticos assintomáticos mostraram uma perda progressiva de função dopaminérgica e 4 desenvolveram DP. Durante o acompanhamento dos pacientes, os níveis de concordância combinadas para disfunção dopaminérgica subclínica e DP foi de 75% em MZ e 22% em DZ. Os autores concluíram por um papel importante da herança genética na DP idiopática. Este resultado pode ser explicado pelo fato de que estudos longitudinais têm maior chance de detectar pares concordantes devido ao acompanhamento extensivo; além disso, a escolha do fenótipo (captação competitiva de 18F-DOPA), um parâmetro único, pode ter aumentado a chance de sucesso de detecção de um efeito real de concordância diferencial.

Cinco anos mais tarde, foi publicado um estudo envolvendo amostra populacional inicial de 14.082 pares de gêmeos suecos em busca de pares afetados por DP (50) e após critérios de inclusão e validação, chegou-se em 247 pares de gêmeos com idade igual ou superior a 50 anos; destes, apenas 2 pares (ambos

dizigóticos) foram concordantes. Segundo os autores este resultado sugere que fatores ambientais específicos para cada indivíduo são mais importantes para a etiologia de DP, no entanto a participação de genes influenciados pelo ambiente e/ou genes de baixa penetrância não pode ser excluída (50).

Em 2008, Thacker e Ascherio (54) realizaram uma meta-análise de dados de agregação familiar em DP com o objetivo de avaliar o RR de DP quando se tem um parente de primeiro grau com a doença. O estudo mostrou que ter um parente de primeiro grau afetado aumenta o risco de ocorrência da doença em 2,9 (IC 95%: 2,2-3,8). Além disso, o RR para pares de irmão foi de 4,4 (IC 95%: 3,1-6,1), enquanto para os pares de pais e filhos foi de 2,7 (IC 95%: 2,0-3,7). Ainda, O RR para o início precoce de PD foi de 4,7 (IC 95%: 3,2-6,8), enquanto que para início tardio de PD foi de 2,7 (IC 95%: 1,9-3,9).

Fazendo uso de informações genéticas mais detalhadas, Hamza e Payami (70) estudaram as famílias de 504 probandos caucasianos (totalizando 2.828 indivíduos) oriundos do *NeuroGenetics Research Consortium* (NGRC) para estimar a herdabilidade de DP. Os probandos foram genotipados para mutações de alto impacto nos genes *PARK2*, *SNCA*, *LRRK2*, *DJ1*, *MAPT* e *NURR1*. Os autores relataram que a herdabilidade *narrow-sense* (h^2) para a amostra populacional total foi de 60% e chamam a atenção para o estudo de Moilanen *et al.* de 2001(71), que obteve resultado semelhante. Em análise posterior, após a exclusão de pacientes portadores de mutações patogênicas mais raras, os mesmos autores obtiveram uma h^2 de 58%, sendo esta leve redução discutida pelos autores com base na frequência destas mutações, que foi de apenas 2% na amostra estudada. Usando um filtro mais restritivo, no qual os autores retiraram da estimativa todas as famílias que possuíam mutações patogênicas ou mutações de alto risco nos genes *SNCA* e *MAPT*, a estimativa caiu para 40%. Os autores discutem este resultado afirmando que ainda falta um significativo efeito genético a ser detectado para o risco de desenvolvimento de DP idiopático, e que estes prováveis genes explicariam 40% da herdabilidade na susceptibilidade à DP. Apesar dos resultados, uma limitação importante de desenho foi assumir que as mutações genotipadas são as únicas de impacto sobre o controle da doença – por exemplo, somente para o gene *PARK2*, haviam mais de 100 mutações já descritas na época do trabalho, que não foram testadas.

Wirdefeldt *et al.*(72) reavaliaram em 2011 os dados do estudo publicado em 2004(50), e adicionaram ao trabalho um desenho de estudo longitudinal para investigar a h^2 de DP. Como resultado, chegou-se a uma herdabilidade de 40% na mesma amostra de um estudo que inicialmente identificou concordância em apenas dois pares de gêmeos DZ. Interessante notar que este estudo encontrou um valor de herdabilidade igual ao estudo de Hamza e Payami (70) após a remoção de algumas variantes de alto impacto, resultado que aponta um efeito genético bastante parecido por trás da DP em amostras populacionais diferentes.

Mais recentemente, um estudo publicado pelo *International Parkinson Disease Genomics Consortium* (IPDGC) (73) combinou dados de genótipo e *imputation* de 5 varreduras de genoma completo (do inglês *genome-wide association studies* - GWAS) anteriores em populações caucasianas (a metanálise alcançou uma amostra populacional de 5.000 casos e 12.000 controles). Os resultados não só confirmam 6 genes associados com DP em estudos anteriores (*SNCA*, *MAPT*, *BST1*, *LRRK2*, *GAK*, e *HLA-DRB5*), como também revelaram outros 5 novos *loci* com evidência de associação. Usando os dados de risco de 11 *loci* PARK, regiões genômicas descritas em ligação à DP, mais os cinco novos apontados pelo estudo, os autores chegaram a um valor de h^2 de 60%, semelhante ao observado em caucasianos no estudo de Hamza e Payami (70).

1.2.2 Modelos de herança para DP

Em 1982, Barbeau e Pourcher(74) sugeriram herança Mendeliana para certos casos de parkinsonismo, em particular aqueles de aparecimento anterior aos 40 anos de idade. Neste grupo, eles verificaram uma incidência de 46% de co-ocorrência de outros casos na família, sendo que na amostra global de DP, essa incidência cai para 10%. Estes autores sugeriram dois subtipos genéticos de DP possíveis: um de herança autossômica recessiva e outro de herança autossômica dominante. Várias análises subsequentes concluíram pela existência de um efeito genético, com forte dependência da idade de manifestação sobre a penetrância do fenótipo (75-79). Por exemplo, um estudo envolvendo 221 pacientes com DP com

idade de aparecimento da doença de 50 anos ou menos, 266 pacientes com idade de aparecimento superior a 50 anos e 409 controles concluiu por um padrão de herança autossômica recessiva entre o grupo de pacientes com aparecimento precoce da doença, mas não no grupo de aparecimento tardio (80).

Poucos estudos examinaram se a agregação familiar de PD é consistente com herança mendeliana usando análise de segregação, estratégia que tem como objetivo determinar o modelo de herança que melhor explica o comportamento de um determinado fenótipo em pedigrees (71, 81-83). Em 2001, Moilanen *et al.* (71) realizaram uma análise de segregação complexa em uma população do norte da Finlândia, utilizando uma amostra populacional de 265 famílias nucleares identificadas através de probandos de casos de DP idiopático e de início precoce. O modelo esporádico foi rejeitado e foi detectada uma heterogeneidade significativa entre as famílias com casos de DP de início precoce e as de início tardio (considerando um idade de corte de 55 anos), com um efeito de gene principal (*major gene*) maior no primeiro grupo do que no segundo. Porém, não foi possível distinguir entre um modelo recessivo com uma penetrância elevada ou um modelo dominante com uma penetrância baixa. No ano seguinte, Maher *et al.* (82) realizaram uma ampla análise de segregação envolvendo 948 casos de DP e seus núcleos familiares, que indicou a presença de um gene principal, de herança Mendeliana, tanto para susceptibilidade à DP quanto para idade de aparecimento da doença.

1.2.3 Estudos de ligação, associação e sequenciamento

Na tentativa de se mapear os genes envolvidos no controle da doença, análises de ligação resultaram na identificação de numerosos *loci* em ligação com DP e seus fenótipos (84-88). Em combinação com estudos de ligação, análise de associação tem sido aplicada na investigação da natureza do componente genético de susceptibilidade à DP (89-91), e como resultado desta estratégia, foram identificadas evidências para associação entre fenótipos da DP e alelos de marcadores de vários genes, como por exemplo, *monoamine oxidase B* (*MAOB*) e

catecol-O-methyltransferase (COMT), genes que codificam enzimas envolvidos no metabolismo da dopamina (92), do gene da interleucina 1-beta (*interleukin 1-beta-IL1B*) (93), do óxido nítrico sintetase (*nitric oxide synthase –NOS*) (94) e o gene *N-acetyltransferase 2* (NAT2) (95), entre outros. Uma mutação funcional no gene da Sinfilina-1 (*synuclein, alpha interacting protein –SNCAIP*) foi encontrada em dois casos inicialmente considerados esporádicos de DP (96).

Análises de associação realizadas através de GWAS detectaram evidências de associação entre novos *loci* e DP; no entanto, por tratar-se de um método indireto, não foi possível a identificação de quais variantes são realmente responsáveis pelo fenótipo (73, 97-107). Com o surgimento de plataformas de sequenciamento de próxima geração (*Next-generation sequencing*, NGS), abordagens de sequenciamento ganharam força, uma vez que tornou-se possível a produção de dados de genoma e/ou exoma completo com custo e tempo reduzidos. Estudos baseados nessa tecnologia frequentemente apontam variantes causais, permitindo um melhor entendimento das bases moleculares dos fenótipos estudados (108, 109). As primeiras iniciativas em DP produziram resultados promissores: dois estudos independentes, publicados simultaneamente, basearam-se na estratégia de sequenciar os exomas de indivíduos informativos, selecionados em poucos pedigrees contendo múltiplos indivíduos afetados por Doença de Parkinson de Início Tardio (DPIT), em busca de variantes raras de alta penetrância associadas à doença. Como resultado, ambos os grupos encontraram, em todos os indivíduos afetados por DPIT, a variante rara c.1858G>A do gene *VPS35* sendo transmitida sob herança dominante. Este gene codifica uma proteína de papel central no complexo retrômero, altamente conservado evolutivamente, importante no sistema de reciclagem de proteínas de membrana e tráfego endossomo-Golgi. Estes estudos não só foram importantes por identificar variantes de alto impacto em DPIT, uma forma de DP que dispunha de poucas informações sobre a influência de fatores genéticos em sua patogênese, como também apontaram uma nova rota celular candidata para estudos em DP (110, 111).

Os resultados apresentados até aqui, dada sua confirmação pela replicação em populações independentes, reforçam o caráter multifatorial e a heterogeneidade genética por trás da doença. Aliás, a observação da existência dos *loci* principais, numerados de PARK1 a PARK18 (TABELA 2) (112), são forte evidência da

heterogeneidade genética na susceptibilidade à DP. Atualmente dos 18 *loci* PARK, 14 deles já tem genes associados (TABELA 2) (112, 113), todos têm função conhecida e muitos participam de processos celulares já descritos como prováveis participantes da patogênese da DP. As exceções são os produtos dos genes *ATP13A2*, *GIGYF2*, *PLA2G6*, *VPS35* e *EIF4G1* que aparentemente não se encaixam nos modelos de patogênese propostos para DP (39).

TABELA 2. *Loci* PARK1 a 18 com suas respectivas localizações cromossômicas e genes associados (quando identificados).

Cromossomo	PARK-Locus	Gene associado	Modelo de herança*	Estudo inicial
4q21-q23	PARK1	<i>SNCA</i>	DP autossômica dominante	Polymeropoulos, 1996 ^a (84)
6q25.2-q27	PARK2	<i>PARK2</i>	DPIP autossômica recessiva	Matsumine, 1997 ^a (114)
2p13	PARK3	Não identificado	DP	Gasser, 1998 ^a (115)
4p14-16.3	PARK4	<i>SNCA</i>	DP autossômica dominante	Farrer, 1999 ^a (116)
4p14	PARK5	<i>UCHL1</i>	DP autossômica dominante	Leroy, 1998 ^b (117)
1p35-p36	PARK6	<i>PINK1</i>	DPIP autossômica recessiva	Valente, 2001 ^a (118)
1p36.33 - p36.12	PARK7	<i>DJ1</i>	DPIP autossômica recessiva	van Duijn, 2001 ^a (119)
12p11.23-q13.11	PARK8	<i>LRRK2</i>	DP autossômica dominante	Funayama, 2002 ^a (120)
1p36	PARK9	<i>ATP13A2</i>	DPIP autossômica recessiva	Hampshire, 2001 ^a (121)
1p32	PARK10	Não identificado	DP	Hicks, 2002 ^a (122)
2q36-q37	PARK11	<i>GIGYF2</i>	DP autossômica dominante	Pankratz, 2003 ^a (87)
Xq21-q25	PARK12	Não identificado	DPIP	Pankratz, 2003 ^a (88)
2p12	PARK13	<i>HTRA2</i>	DP autossômica dominante	Strauss, 2005 ^b (123)
18q11	PARK14	<i>PLA2G6</i>	DPIP autossômica recessiva	Gao, 2009 ^a (124)
22q12-q13	PARK15	<i>FBXO7</i>	DPIP autossômica recessiva	Shojaee, 2008 ^a (125)
1q32	PARK16	Não identificado	DP	Satake, 2009 ^c (104)
16q12	PARK17	<i>VPS35</i>	DP autossômica dominante	Zimprich, 2011 ^d (110), Vilarino-Guell, 2011 ^d (111)
3q27	PARK18	<i>EIF4G1</i>	DP autossômica dominante	Chartier-Harlin, 2011 ^d (126)

*DP: Doença de Parkinson; DPIP: Doença de Parkinson de Início Precoce. a: Estudos de ligação; b: Estudos de associação; c: GWAS; d: Sequenciamento de exoma completo. Todos os estudos citados

são referentes à descrição inicial do *locus* encontrado, independentemente da identificação de gene.
 FONTE: Adaptado de Lubbe, S. e Morris, H.R., 2013 (112) e PDgene database(127).

Dos genes descritos na TABELA 2, os que tiveram associação com fenótipos de DP consistentemente replicados em populações diferentes são *SNCA*, *PARK2*, *PINK1*, *DJ1* e *LRRK2*. Hoje, há consenso de que a ocorrência de mutações já descritas e prováveis novas mutações de alto impacto funcional em um destes cinco genes leva a um aumento do risco de desenvolvimento de DP (5). Mutações de ganho de função nos genes *SNCA* e *LRRK2* causam DP autossômica dominante, enquanto que mutações de perda de função nos genes *PARK2*, *PINK1* e *DJ1* causam DP autossômica recessiva. Ainda, sabe-se que polimorfismos comuns no *SNCA* e *LRRK2* exercem pequeno mas significativo efeito (OR de 1.20 a 1.5) no risco de desenvolvimento de formas não Mendelianas de DP. Estimativas apontam que as formas monogênicas de DP correspondem a aproximadamente 30% de casos familiares e aproximadamente 5% de casos esporádicos (128, 129).

1.3 UMA NOVA PERSPECTIVA SOBRE A REGIÃO 6q25-q27: *PARK2* É O ÚNICO RESPONSÁVEL PELA LIGAÇÃO DO LOCUS *PARK2* À DOENÇA DE PARKINSON DE INÍCIO PRECOCE?

Em 1996, foram descritas 12 famílias japonesas apresentando 17 casos de DP de início precoce (DPIP). Com base em análise de heredograma, com vários casos de consanguinidade, e dados de características clínicas, os autores concluíram que a transmissão da doença seguia um padrão de herança autossômica recessiva, e que a entidade fisiopatológica, denominada Doença de Parkinson de Início Precoce Autossômico-Recessivo (DPIP-AR) era distinta das outras formas de DPIP (130). Subsequentemente, uma análise de ligação envolvendo 14 famílias contendo membros portadores de DPIP-AR revelou forte ligação (LOD score=7,71) entre a doença e a região cromossômica 6q25-q27 (114). Neste estudo, um dos marcadores utilizados (D6S253) estava deletado em um dos pacientes. Em artigo complementar, o mesmo grupo de pesquisadores descreveu um novo gene localizado na região 6q25-q27, denominado por eles *PARKIN*, que continha a

microdeleção do marcador D6S253. Mutações neste novo gene foram encontradas em quatro outros pacientes de famílias não relacionadas. Os autores concluíram que mutações no gene *PARKIN* (hoje *PARK2*) seriam responsáveis pela patogênese do DPIP-AR nestes pacientes (131).

Dados acumulados desde então indicam que mutações no gene *PARK2* respondem por grande parte dos casos de DPIP-AR (12, 132, 133). Estas mutações apresentam-se de forma muito variada, incluindo várias mutações de ponto e rearranjos envolvendo todos os 12 exons do gene (12, 22, 133, 134). A frequência destas mutações já foi estimada em 50% de famílias com DPIP-AR e 15-18% de pacientes com DPIP isolado (12, 13).

Resultados de uma análise de ligação, conduzida em 174 famílias com múltiplos membros afetados por DP idiopática, identificou ligação ($LOD\ score = 5.47$) entre a doença na sua forma de aparecimento anterior aos 40 anos de idade e marcadores localizados no gene *PARK2* (135). West *et al.* (2002) (134) encontraram evidência para associação entre polimorfismo localizado na região promotora do gene *PARK2* com DP idiopática em uma população de 296 casos e 184 controles ($OR = 1,52$). Funcionalmente, o gene *PARK2* codifica uma enzima denominada parkina, uma E3 proteína-ubiquitina ligase responsável pela ubiquitinação de proteínas-alvo, dirigindo-as para degradação no proteassomo (136, 137).

A fim de se avançar no entendimento da relação entre mutações do gene *PARK2* e parkinsonismo, Perez e Palmiter (2005) (138) desenvolveram, por deleção dirigida do exon 2, uma linhagem de camundongo nocaute para *Park2*. Surpreendentemente, esta linhagem deficiente para parkina não diferiu da linhagem controle em relação a um extenso protocolo de avaliação de perfil comportamental, que incluía funções neurológica, motora, emocional, de aprendizado e memória. Os autores concluíram que camundongos deficientes para parkina não são um modelo adequado para parkinsonismo. Por outro lado, uma linhagem amplamente conhecida de camundongo, denominada *Quaking*^{via} (Qk^v), vem sendo utilizada como modelo para estudo de desmielinização de sistema nervoso central. Estes animais são portadores de uma mutação espontânea, recessiva, do gene *Quaking* (*Qki*) localizado no cromossomo 17, cujo dano na sequência de nucleotídeos consiste em uma deleção de extensão de 1 Mega base (Mb) (139).

Camundongos qk^v homozigotos (qk^v/qk^v) são animais viáveis, mas apresentam 5 a 10% da quantidade normal de mielina no sistema nervoso central (SNC). Consequentemente, por volta de 10 dias após o nascimento, estes animais desenvolvem tremores vigorosos (140), um dos fenótipos presentes na síndrome parkinsoniana. Interessantemente, foi demonstrado pelo trabalho de Lockhart *et al.* (141) em 2004, que a deleção qk^v afeta grande parte da sequência codificante dos genes vizinhos *Park2* e *Pacrg*, e estende-se por aproximadamente 900 pares de bases (pb) a montante do sítio de início de transcrição do gene *Qki*. Paralelamente ao estudo de Lockhart, trabalho independente de Lorenzetti *et al.* (2004) (142) demonstrou que animais da linhagem Qk^v são deficientes em parquína, resultados consistentes com outros estudos que demonstraram a importante deleção localizada no gene responsável por sua codificação.

Pouco tempo após a publicação destes trabalhos, o *International Radiation Hybrid Mapping Consortium* mapeou um gene homólogo ao *Qki* em humanos, denominado *QUAKING HOMOLOG OF MOUSE (QKI)*, no cromossomo 6q25-q27, em posição homóloga àquela observada no cromossomo 17A do camundongo, ou seja, adjacente aos genes *PARK2* e *PACRG* (FIGURA 1A e B). Dado interessante de expressão gênica revelou que o gene *QKI* está menos expresso em amostras de tecido cerebral obtido na autópsia de pacientes portadores de esquizofrenia (143). Embora as desordens sejam clínica e fisiopatologicamente distintas, implicação do gene *QKI* em esquizofrenia, também uma doença do sistema nervoso central, pode levantar suspeitas sobre a participação do gene *QKI* na patogênese da DPIP.

O gene humano *QKI* tem quatro variantes diferentes geradas por *splicing* alternativo (FIGURA 1C) e codifica a proteína QKI da família de domínios KH, cuja função é identificar e ligar-se a RNA (143). Ainda, evidências apontam que QKI tem participação na estabilização de pré-RNAm e sua exportação (144).

As isoformas do *QKI* e seus padrões gerais de expressão conhecidos são o QKI-5, altamente expressa no cérebro fetal, QKI-6, QKI-7a e QKI-7b, que têm suas expressões aumentadas durante o período de mielinização (145). Estas isoformas exibem diferentes padrões de localização subcelular e de expressão programada, características necessárias às suas funções sofisticadas e, às vezes, até opostas. Muitos RNAm contêm elementos de reconhecimento da proteína QKI que podem ligar-se diretamente ao QKI *in vitro*. Estudos de bioinformática identificaram diversos

RNA_m alvos do QKI, pertencentes às classes de função celular de especificação de destino celular, diferenciação/crescimento, migração e comunicação célula-célula (140).

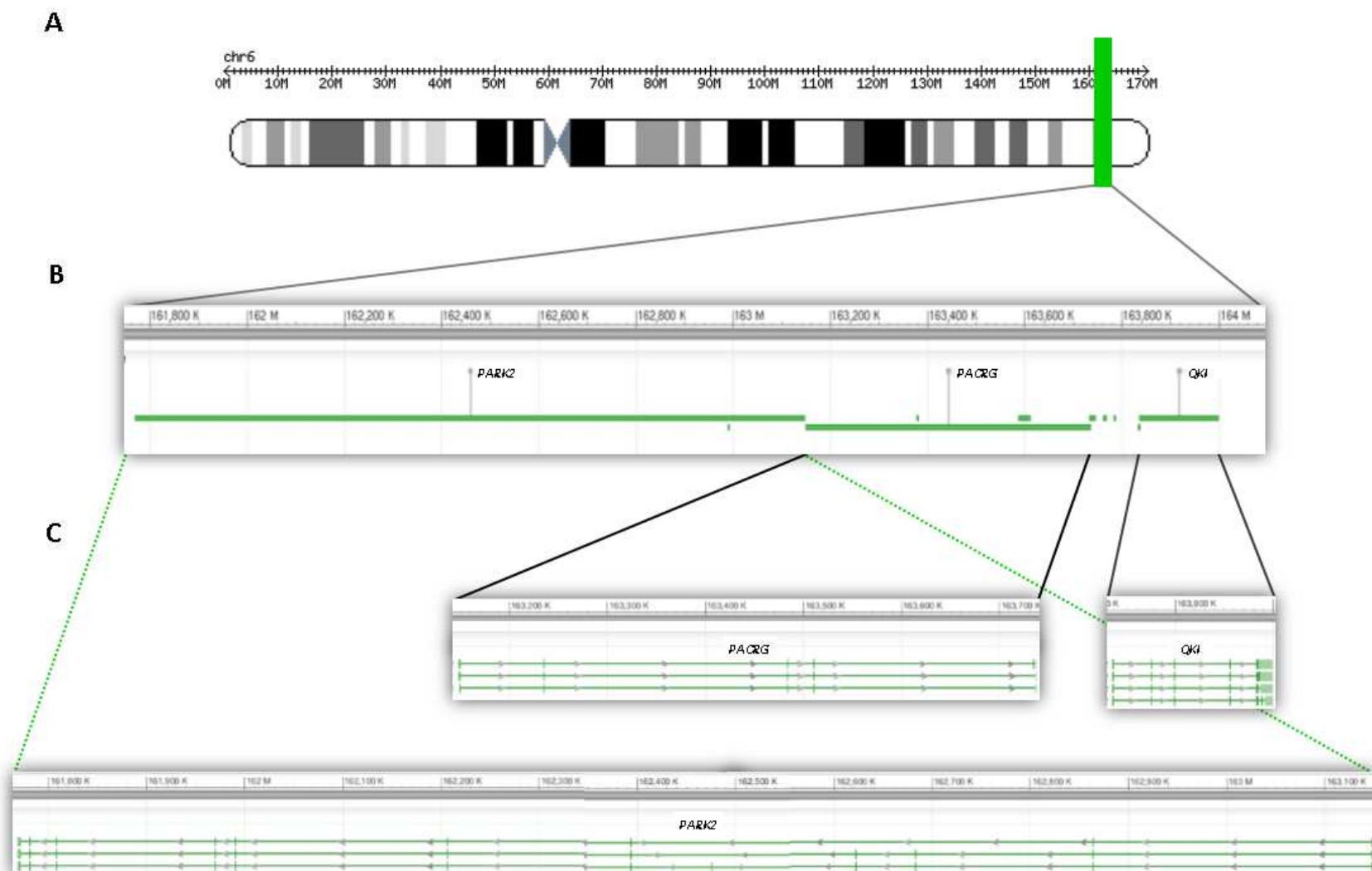


FIGURA 1. Representação de uma região de aproximadamente 2,5Mb no cromossomo 6q humano. **A)** Localização no cromossomo. **B)** Organização genômica dos genes *PARK2*, *PACRG* e *QKI*.
 FONTE: Adaptado de NCBI (146, 147), Ensembl (148-150) e HapMap (151, 152).

Interessantemente, um estudo de Backx *et al.* (2010) (153) descreveu uma paciente que exibia características clínicas de uma desordem denominada “síndrome do cromossomo 6q”, cuja alteração consiste em uma deleção submicroscópica subtelomérica de um segmento do cromossomo 6 que desorganiza a estrutura do gene *QKI*, mas esta alteração não gera fenótipos compatíveis com parkinsonismo como observado em camundongos. Nesta paciente, a mutação levou a uma redução de 50% na expressão do gene *QKI* sem o comprometimento de outras regiões genômicas, indicando que a insuficiência da proteína QKI está envolvida em uma parcela significativa da síndrome da deleção do 6q. De maneira geral, as características fenotípicas principais desta síndrome são deficiências cognitivas, hipotonia, convulsões, anormalidades cerebrais e características dismórficas como pescoço curto, nariz largo, orelhas largas abaixo da posição comum e cantos da boca voltados para baixo. Contrastando com o fenótipo observado em camundongos *qk^v*, análise por ressonância magnética do cérebro da paciente não evidenciou mielinização anormal, nem mesmo anormalidades estruturais. Ainda, até o momento da realização daquele estudo, esta paciente não desenvolveu quadros de epilepsia. Apesar da estrutura cerebral aparentemente normal, este indivíduo possuía leve grau de deficiência cognitiva. Estas situações descritas acima dão consistência à hipótese de que, provavelmente, a haplosuficiência deste gene é suficiente para manter as funções de mielinização.

Outro gene que merece investigação conjunta ao *PARK2* e *QKI* é o *PACRG*, uma vez que: (i) a organização genômica do *locus* 6q25-27 em humanos e a região homóloga em camundongos posiciona o *PACRG* muito próximo de *PARK2*, entre este e *QKI*; (ii) em camundongos *qk^v* há uma deleção comprometendo os 3 genes da região; (iii) a arquitetura dos genes *PARK2/PACRG*, coloca os dois genes sob o controle de um promotor bidirecional (FIGURA 1B)(154), e (iv) o produto protéico do *PACRG* também faz parte do corpúsculo de Lewy (155, 156). Apesar de ser um bom candidato posicional, e da evidência de *PACRG* em inclusões intracelulares, ao investigar a participação do gene *PACRG* na DP, um trabalho apontou ausência de associação deste gene na amostra populacional estudada. No entanto, os autores não excluem o gene como participante das vias que desencadeiam a desordem (156). Estudo de Imai *et al.* realizado em 2003 (157) fortalece esta hipótese, uma vez que sugeriu que *PACRG* pode funcionar como um mediador da

morte celular neuronal. Postula-se que o produto gênico de *PACRG* participa ou interage com parkina na via de degradação proteassomal; no entanto o significado biológico desta interação ainda não é compreendido.

Diante do exposto e resgatando a informação de que variações de *PARK2* respondem por aproximadamente 50% dos casos de DPIP-AR e até 18% dos casos de DPIP esporádico (12, 13), é razoável supor que mutações em *PACRG* e *QKI*, genes em estreita proximidade e de envolvimento em condições neuropatológicas comuns, podem ser coadjuvantes ou tão importantes quanto variações de *PARK2* na fisiopatologia de DPIP.

1.4 DESCRIÇÃO DA ESTRUTURA E FUNÇÕES DOS GENES EM ESTUDO

1.4.1 *PARK2*

Nomenclatura oficial:	parkinson protein 2, E3 ubiquitin protein ligase (parkin)
Localização:	6q25.2-q27
Início:	161.768.452 pb do pter
Fim:	163.148.834 pb do pter
Tamanho:	1.380.383 bases
Orientação:	Cadeia reversa
RefSeq:	NC_000006.11

Este gene tem extensão total de 1,38 Mb e sua estrutura genômica possui 12 exons (FIGURA 1C) que dão origem a 7 isoformas de RNAm. Destas isoformas, 3 têm produto protéico caracterizado e registrado em bancos de dados (FIGURA 1C), outras 3 têm estrutura compatível com expressão mas ainda não têm registro de produtos de tradução e a última não é expressa devido ao mecanismo de decaimento do RNAm (*nonsense mediated decay*), um processo que detecta mutações *nonsense* e previne que proteínas truncadas ou aberrantes sejam expressas (147, 148, 158).

A proteína codificada pelo *PARK2*, parkina, faz parte do complexo multiprotéico E3 ubiquitina ligase e está envolvida na degradação de proteínas pela via do sistema ubiquitina-proteassomo e na mitofagia (136). A proteína PINK1 também está envolvida neste processo e é responsável pelo recrutamento da parkina para mitocôndrias danificadas (159).

1.4.2 *PACRG*

Nomenclatura oficial:	PARK2 co-regulated gene
Localização:	6q26
Início:	163.148.164 pb do pter
Fim:	163.736.524 pb do pter
Tamanho:	588.361 bases
Orientação:	Cadeia direta
RefSeq:	NC_000006.11

Este gene tem extensão total de 588Kb e sua estrutura genômica possui 7 exons (FIGURA 1C). Variantes de transcritos que codificam para 8 isoformas diferentes foram encontrados para este gene. Destas isoformas, 3 têm produto protéico caracterizado e registrado em bancos de dados (FIGURA 1C), outras 2 têm estrutura compatível com expressão mas ainda não têm registro de produtos de tradução, 1 não é expressa devido ao mecanismo de decaimento do RNAm, e 3 são transcritos não codificantes de proteínas (duas delas não contêm o início da janela de leitura e uma contém uma sequência intrônica) (147, 148, 158).

PACRG codifica uma proteína conservada entre metazoários. Nos vertebrados, este gene está ligado em um arranjo de cabeça-a-cabeça com o *PARK2* adjacente (FIGURA 1B). Estes genes são co-regulados em diversos tecidos e compartilham um promotor bidirecional (154). A proteína codificada pelo *PACRG* forma um grande complexo molecular com chaperonas, incluindo as proteínas *heat shock* 70 e 90 e os componentes de chaperonina. Esta proteína é também um componente de corpos de Lewy em pacientes com doença de Parkinson (155, 156), e reprime a morte de células neuronais induzida por receptores Pael deformados

(157).

1.4.3 QKI

Nomenclatura oficial:	QKI, KH domain containing, RNA binding
Localização:	6q26
Início:	163.835.032 pb do pter
Fim:	163.999.628 pb do pter
Tamanho:	164.597 bases
Orientação:	Cadeia direta
RefSeq:	NC_000006.11

Este gene tem extensão total de 164Kb e sua estrutura genômica possui 8 exons (FIGURA 1C). Até o momento, 17 isoformas de RNAm já foram encontrados. Destas isoformas, 4 têm produto protéico caracterizado e registrado em bancos de dados (FIGURA 1C), 8 têm estrutura compatível com expressão mas ainda não têm registro de produtos de tradução, outras 3 não são traduzida devido ao mecanismo de decaimento do RNAm e as 2 últimas têm seu processamento interrompido por possuírem introns em suas sequências (147, 148, 158).

A proteína codificada por este gene é uma proteína de ligação ao RNA (*RNA-binding protein*) que regula o *splicing* do pré-RNAm, a exportação de RNAm do núcleo, a tradução protéica, e a estabilidade do RNAm. A proteína QKI está envolvida na mielinização e diferenciação oligodendrocítica e pode desempenhar um papel no controle da ocorrência da esquizofrenia (144).

1.5 APLICAÇÃO DE PLATAFORMA DE SEQUENCIAMENTO DE PRÓXIMA GERAÇÃO PARA A IDENTIFICAÇÃO DE VARIANTES CAUSAIS

A era dos GWAS foi marcada pela expectativa de que, ao se utilizar polimorfismos comuns para cobrir genomas completos em estudos de associação, as bases moleculares dos mais variados fenótipos complexos seriam finalmente

desvendadas. No entanto, com a popularização desta abordagem, vem se tornando cada vez mais claro que grande parte do efeito genético controlando susceptibilidade a doenças comuns continuava por ser detectada (160): com raras exceções, mais de 90-95% do componente genético de várias doenças complexas continuou oculto mesmo após extensiva análise por GWAS, fenômeno que deu origem ao termo “herdabilidade oculta” (*hidden heritability*) (160, 161). Ao mesmo tempo que se reconhece que variantes comuns não são suficientes para explicar doenças comuns, as plataformas de sequenciamento de próxima geração, instrumentos que permitem a produção de sequências de exomas ou mesmo genomas completos, ganham destaque por sua capacidade de apontar, com grande precisão, variantes causais dos mais variados fenótipos; como por exemplo, a identificação da base genética de doenças raras como a doença de Charcot–Marie–Tooth (162), as síndromes de Schinzel-Giedion (163), de Fowler (164), de Miller (165), a perda auditiva não sindrômica (166) e o sarcoma de Kaposi (167). Neste cenário, nasceu a hipótese “Doença Comum – Múltiplas Variantes Raras (DCMVR)”, baseada na premissa de que múltiplas variações raras, com penetrância relativamente alta, podem explicar parte da herança oculta observada nos estudos de susceptibilidade genética a doenças comuns (168, 169).

A tecnologia NGS tem revolucionado a nossa compreensão de como genes ou regiões genômicas estão envolvidos na patogênese de doenças humanas, mesmo naquelas sem um claro componente genético, por permitir uma exploração livre de hipótese de grandes extensões ou mesmo de todo o genoma (170). Vale ressaltar uma diferença marcante entre resultados obtidos de GWAS ou NGS: GWAS faz uso de marcadores que têm a função de representar variações causais indiretamente, enquanto que NGS pode identificar diretamente variantes causais. Esta é a vantagem principal de abordagens de sequenciamento em relação a painéis de genotipagem padronizados, que ganha cada vez mais destaque uma vez que a hipótese DCMVR tem se consolidado entre doenças comuns (110, 111, 171).

Neste trabalho, o foco é a aplicação da plataforma *Ion PGM* no sequenciamento de alta cobertura de regiões exônicas dos genes *PARK2*, *PACRG* e *QKI* em busca de variantes de nucleotídeo único (*Single nucleotide variant* - SNV) causais da DPIP. A plataforma *Ion PGM* emprega o método NGS de sequenciamento conhecido como “sequenciamento por síntese”, através do qual o sequenciamento é

realizado em tempo real e é baseado na detecção de íons de hidrogênio gerados durante a reação de polimerização de fitas de DNA complementar à um alvo de interesse.

1.5.1 Bases conceituais: plataforma *Ion Torrent PGM* e características de dados NGS

1.5.1.1 Biblioteca de DNA

Biblioteca de DNA é o conjunto de moléculas-alvo preparadas a partir de um material de partida (DNA genômico, regiões de interesse obtidas por captura ou por amplificação em reação PCR, entre outros), de tal forma que este material tenha concentração e modificações específicas que permitirão a etapa de preparo dos moldes para sequenciamento (*Template preparation*). No preparo da biblioteca, todas as moléculas-alvo passam por uma reação de ligação à adaptadores universais (A e P1) em ambas as extremidades. Estes adaptadores são sequências que hibridam com *primers* para que todas as moléculas-alvo sejam amplificadas de maneira indiscriminada em uma PCR convencional. Ao término desta etapa tem-se, além dos alvos inicial, clones que também tem sequência de adaptadores em suas extremidades. Este adaptadores serão importantes para a próxima etapa (108, 109, 172).

1.5.1.2 PCR em emulsão (ePCR)

Esta reação de PCR em condições especiais é a etapa chave do processo de preparo dos moldes (*Template preparation*), e tem o objetivo de fixar as moléculas da biblioteca de DNA em microesferas (*beads*, ou *Ion Sphere Particles* - ISP), o suporte para a reação de sequenciamento. Para isso, a biblioteca passa por uma nova rodada de amplificação – porém, desta vez, o meio de reação é uma emulsão

água-óleo. É preparada uma solução aquosa contendo a biblioteca, DNA polimerase e deoxinucleotídeos, em presença de microesferas que tem um *primer* universal ligado à sua superfície. O meio aquoso preparado é emulsionado com uma fase oleosa de modo que, em condições ideais, a emulsão gere uma gotícula de água com uma microesfera, polimerases e apenas um fragmento de DNA, unidade denominada “microreator”. Idealmente, em um tubo de reação, tem-se milhares de microreatores, que quando submetidos a ciclos de PCR, farão com que os adaptadores da biblioteca hibridem com o *primer* da microesfera, produzindo novas cadeias de DNA fita simples ligadas às ISP. Ao final da reação, a emulsão é quebrada e as microesferas carregadas são recuperadas e separadas daquelas onde não houve amplificação. A microesfera ideal é aquela que apresenta apenas um tipo de amplicon a partir do original, dita microesferaclonal. Estas serão processadas de acordo com o protocolo de preparo dos moldes para adequação à reação de sequenciamento (108, 172).

1.5.1.3 Princípio do sequenciamento pela plataforma *Ion Torrent PGM*

As microesfera carregadas são depositadas em chip semicondutor que contém em sua superfície milhares de micropoços, cada um com diâmetro suficiente para acomodar apenas uma delas. Abaixo dos micropoços, existe uma placa-sensor que é sensível a alterações de pH. O sequenciamento ocorre quando o chip é preenchido com uma solução contendo apenas um tipo de deoxinucleotídeo não modificado. Há uma ordem de fluxo de deoxinucleotídeos (*Flow order*) e, a cada fluxo, se a base injetada no sistema for complementar ao molde, a polimerase incorpora o deoxinucleotídeo à cadeia de DNA nascente em um número correspondente de vezes que houver pareamento com a fita alvo. Cada incorporação de nucleotídeos é acompanhada da liberação de um íon H^+ que induz alteração de pH no poço de reação; esta alteração é detectada pela placa sensor ISFET (*Ion-Sensitive field-effect transistor*). As alterações de pH são transformadas em impulsos elétricos que são interpretado por algoritmos especializado. De acordo com a ordem de bases injetadas no sistema e a magnitude de alteração de pH (que é proporcional ao número de bases incorporadas) detectadas, a plataforma *Ion PGM*

constrói sequências curtas com média de 200pb para cada poço, denominadas *reads* (109, 173).

Este método de sequenciamento por semicondução é uma variante da tecnologia de pirosequenciamento; assim, apesar da entidade de detecção ser íons próton, o método apresenta uma limitação em comum com o pirosequenciamento: a baixa resolução de sequências homopoliméricas, especialmente acima de 7 unidades de polímero (mer). Após a incorporação do primeiro deoxinucleotídeo em uma região homopolimérica, o sinal detectado não aumenta linearmente a cada nova incorporação, e quanto mais longa a sequência, menor a resolução. Apesar desta limitação, a plataforma *Ion PGM* tem três aspectos que a tornam atrativa: (i) pode gerar até 1 gigabase de dados brutos, que correspondem a aproximadamente 550.0000 *reads*; (ii) é baseada em um sistema não óptico de aquisição de dados, e no uso de nucleotídeos não modificados; (iii) a tecnologia de chip semicondutor torna o processo extremamente rápido, com um tempo de sequenciamento de apenas 2 horas (109, 173).

1.5.1.4 Características dos dados NGS

Os dados obtidos a partir de plataforma de NGS diferem daqueles obtidos pelo método clássico de Sanger, tanto no volume quanto no formato de saída. Uma sequência alvo de 500 pb submetida a sequenciamento clássico gera apenas uma sequência linear conhecida como eletroferograma. Esta sequência é facilmente mapeada contra uma sequência referência, sem grande necessidade de recursos computacionais. Um mesmo alvo submetido a NGS na plataforma *Ion PGM* gera (a exemplo deste trabalho) no mínimo 3.000 sequências curtas (*reads*). *Reads* são sequências de extensão variável que têm, em média, 200 pb (no caso da plataforma *Ion PGM*); limitada não pelo tamanho da molécula alvo, mas sim, pela química empregada (FIGURA 2). *Reads* não clonais são aquelas que têm origem e término em diferentes pontos da sequência referência, e são as que efetivamente são consideradas para a cobertura de uma base. Cobertura de bases se refere a quantas *reads* não clonais alinharam com a base referência (FIGURA 2) (109, 174).

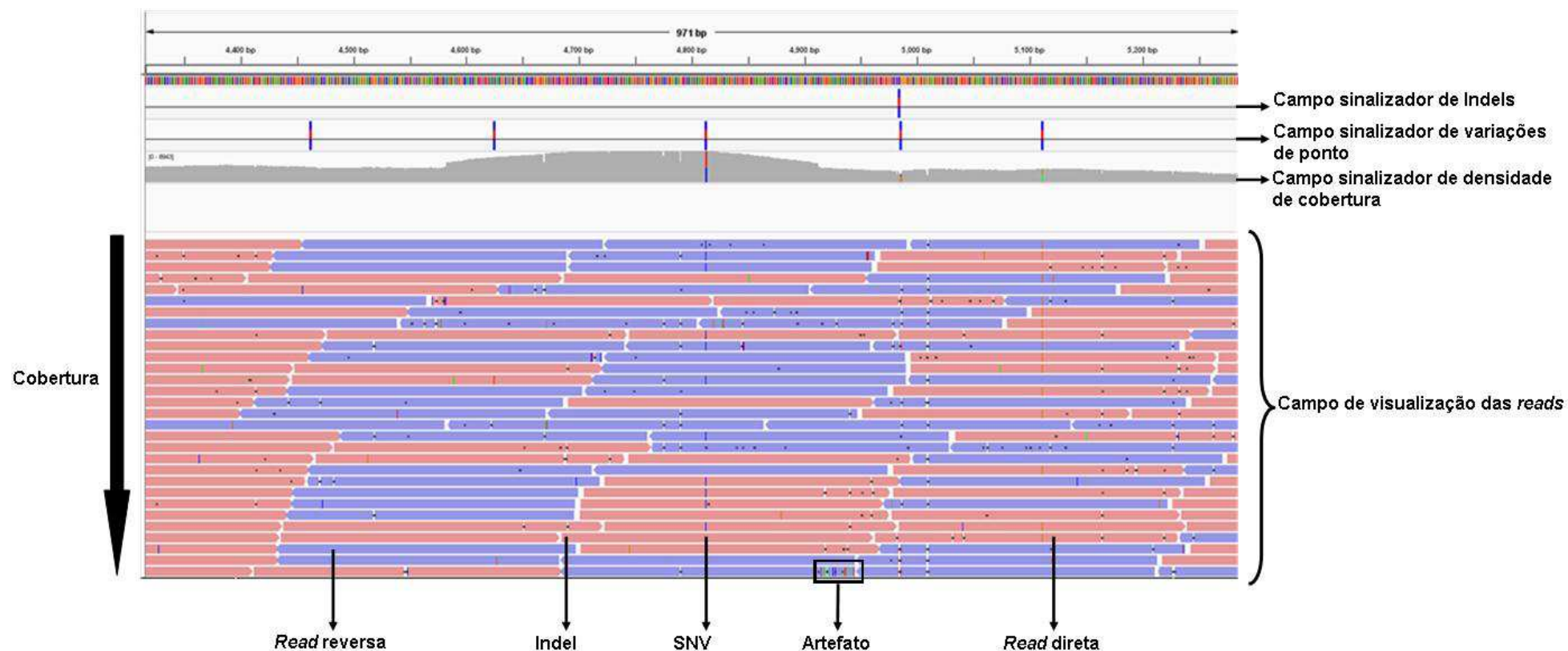


FIGURA 2. Visualização de alinhamento de *reads* e apresentação dos principais indicadores, no programa IGV. Em rosa destacam-se as *reads* diretas e em azul, as *reads* reversas. Os círculos pretos nas *reads* representam Indels e as linhas verticais nas cores verde, azul, amarelo e vermelho representam SNV. Código de cores: A, C, G, T. Indel: Inserção/deleção; SNV: *Single Nucleotide Variant*.

Em virtude do imenso volume de dados produzidos por plataformas NGS, gerou-se a necessidade de criar arquivos com codificações especiais, que contivessem todo o volume sequências e seus parâmetros de qualidade. Estes arquivos são conhecidos como SAM (*Sequence Alignment/Map*), sendo mais utilizada a sua versão binária BAM, que comporta o mesmo volume de dados em um menor espaço. Mesmo fazendo uso de um arquivo extremamente compacto – mas que ainda contém grande volume de dados –, a demanda computacional para se mapear e analisar estas sequências é consideravelmente maior (109, 174).

Outra característica que faz a análise dos dados de NGS mais exigente quanto a demanda computacional é a necessidade de aplicação de algoritmos e protocolos de avaliação de qualidade e veracidade dos dados gerados. Essa exigência tem como base o fato de que plataformas NGS são mais propensas a erros de sequenciamento em comparação com o método de Sanger. Desta forma, essas plataformas são excelentes ferramentas para análise de varredura, no entanto seus achados devem ser validados por plataforma diferente, sendo as de sequenciamento clássico as de escolha (175).

1.5.2 Estratégia de *pooling* para detecção de variantes

O desenvolvimento e aprimoramento de plataformas NGS representaram importante avanço tanto por tornar mais viável a busca de variantes de relevância para o desenvolvimento dos mais variados fenótipos, como também por oferecer a possibilidade de sequenciar grande quantidade de regiões genômicas por custo acessível. No entanto, a aplicação destas plataformas no sequenciamento de um grande número de indivíduos ainda é economicamente limitado na maioria dos laboratórios. Neste cenário, a estratégia de *pooling* vem sendo adotada, por permitir, dependendo do método empregado, a avaliação de variações genéticas em um grupo de pessoas pelo custo virtual de um indivíduo (176-178).

Dois métodos de *pooling* podem ser utilizados, o primeiro baseia-se em um preparo de biblioteca com *barcodes* (sequência conhecidas e padronizadas de DNA que servem como identificadores quando ligados ao DNA-alvo). Este método envolve um protocolo de preparo individual de biblioteca para cada amostra, que ao

final gera sequências identificadoras únicas ligadas à todas as moléculas de DNA de um indivíduo, ou seja: cada indivíduo terá um *barcode* diferente ligado à sua amostra de DNA. Este conjunto de amostras é utilizado no preparo de uma única solução contendo DNA de todos os indivíduos que se deseja sequenciar. A partir deste *pool*, a solução é tratada como uma única amostra até o término da etapa de sequenciamento. Na análise, as sequências *barcode* permitem a individualização das amostras (109, 179).

O segundo método de *pooling* consiste no preparo da solução de amostras de DNA de todos os indivíduos que se deseja sequenciar, sem a identificação de cada componente. Este método é uma abordagem de amostragem populacional, que visa avaliar a posição de variantes no DNA sem se preocupar em distinguir, em uma etapa inicial, os indivíduos portadores das variantes encontradas (176-178).

Como usual, ambas as técnicas têm vantagens de desvantagens, sendo a desvantagem para o primeiro método o custo mais alto em comparação com o segundo método, e uma maior demanda de bioinformática para a análise das sequências. Já o método de *pooling* sem *barcodes* exige menor demanda bioinformática e torna o custo de sequenciamento tão baixo quanto para um indivíduo; porém, tem a óbvia desvantagem de dificultar a diferenciação entre variações reais, principalmente as de menor frequência, e erros de sequenciamento, que compromete uma estimativa precisa de frequências alélicas e limita a exploração dos dados quanto a possíveis consequências individuais (176-178).

2 JUSTIFICATIVA

Este trabalho tem dois aspectos distintos, mas complementares: primeiro, este é um estudo clássico de investigação científica baseada em hipótese; segundo, esta hipótese foi investigada utilizando-se uma nova tecnologia, cuja aplicação foi, *per se*, um desafio científico e tecnológico.

Considerando-se que: (a) mutações no gene *PARK2* não explicam mais de 50% dos casos de DPIP; (b) Mutações em *PARK2* estão associadas à uma pequena parcela dos casos de DP idiopática; (c) modelo animal nocaute para *Park2* não apresenta fenótipo de parkinsonismo; (d) modelo animal *Qk^v* portador de mutação no gene *Qki* apresenta fenótipo compatível com parkinsonismo; (e) a linhagem de camundongo *Qk^v* é nocaute duplo para os genes *Park2* e *Pacrg* e apresenta dano na estrutura do *Qki*; e (f) o gene *QKI* está localizado adjacente aos genes *PARK2* e *PACRG* no cromossomo humano 6q25-q27, em homologia ao cromossomo 17A do camundongo; nossa hipótese é que mutações nos genes *QKI* e/ou *PACRG*, em conjunto ou não com mutações de *PARK2*, podem ser necessárias ou suficientes para o desenvolvimento de doença DPIP em seres humanos.

Esta hipótese foi testada utilizando-se técnica de sequenciamento de alta cobertura, de próxima geração. A aplicação desta nova tecnologia levou ao desenvolvimento e/ou adaptação de estratégias de produção e análise de dados que apresentaram, em si, um desafio científico. A descrição cuidadosa dos procedimentos, bem como dos resultados dos experimentos e técnicas, com suas limitações e aplicações, pode contribuir para acelerar a utilização desta poderosa tecnologia em estudos futuros.

3 OBJETIVOS

O objetivo geral deste estudo foi sequenciar as regiões codificantes dos genes *QKI*, *PACRG* e *PARK2* de indivíduos portadores de DPIP e controles não afetados em busca de variações genéticas que possam predispor indivíduos à DPIP.

Especificamente, este trabalho buscou:

1. Sequenciar regiões codificantes dos genes *QKI*, *PACRG* e *PARK2* nos *pools* dos indivíduos recrutados;
2. Estabelecer e aplicar protocolos de rastreamento e filtragem das variantes genéticas encontradas;
3. Detectar variações nos genes que possam explicar conjunta ou separadamente o fenótipo de DPIP;
4. Inferir o impacto funcional previsto como consequência da ocorrência das mutações detectadas.

4 MATERIAIS E MÉTODOS

4.1 INSTITUIÇÃO DE ORIGEM E INSTITUIÇÕES PARCEIRAS

Este foi um estudo colaborativo, sendo a unidade/instituição de origem o Departamento de Genética da Universidade Federal do Paraná (UFPR), e as instituições parceiras, o Hospital de Clínicas da UFPR (HC), a Associação Paranaense de Portadores de Parkinson (APPP) – ambas colaboradoras na etapa de busca de pacientes e coleta do material biológico –, a Pontifícia Universidade Católica do Paraná (PUCPR) e o Instituto Carlos Chagas (ICC) – instituições que foram sede para a realização dos experimentos e análises de dados.

4.2 DESENHO DE ESTUDO

Um resumo do desenho geral do estudo está apresentado, na forma de fluxograma, na figura 3. Em seguida, é apresentada uma descrição de cada etapa, antecedida, para fins didáticos, de um breve relato dos seus objetivos, quando pertinente – em especial, quando o procedimento é parte específico do método de sequenciamento empregado.

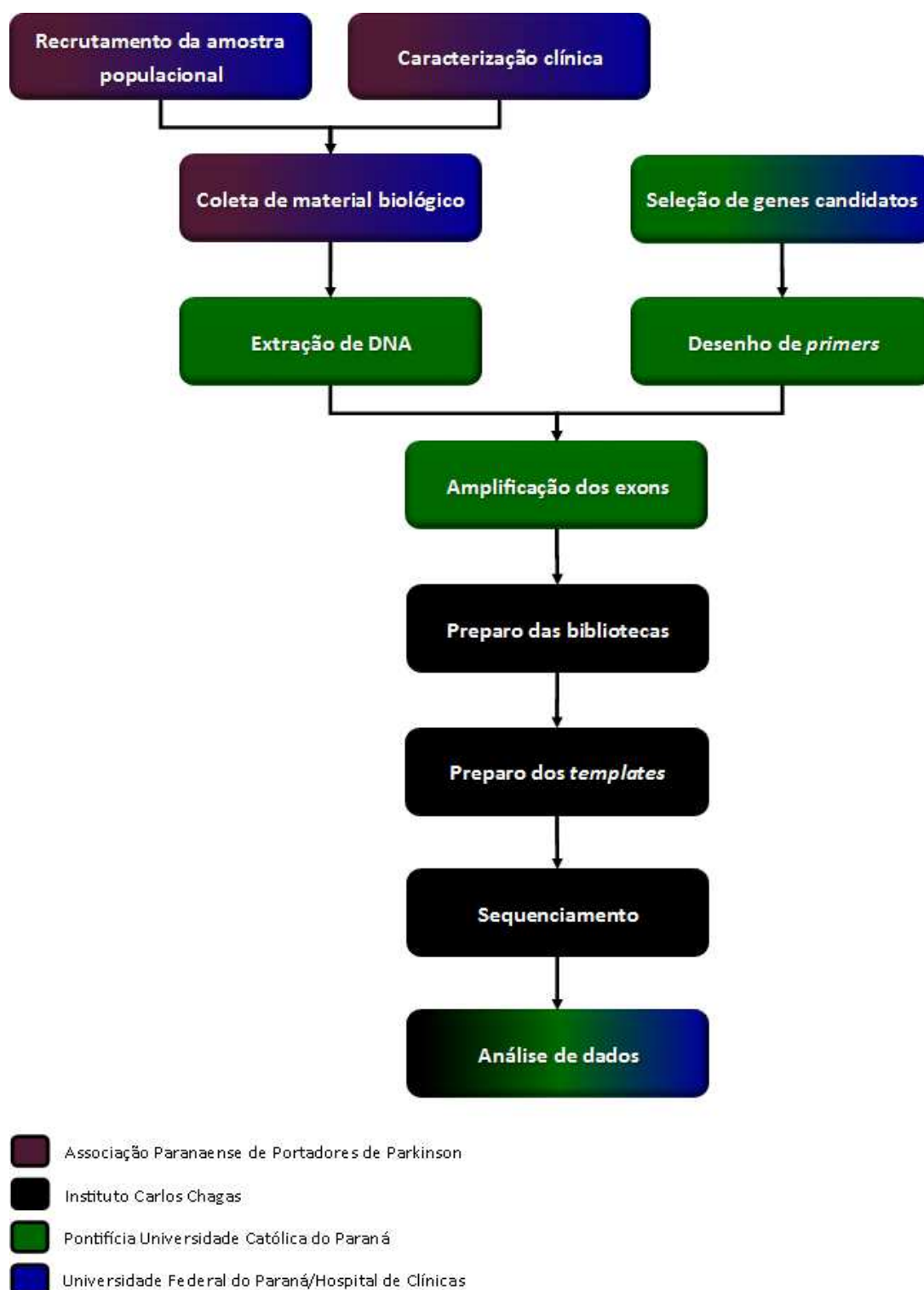


FIGURA 3. Desenho geral do estudo. As cores indicam a(s) instituição(ões) envolvida(s) em cada etapa.

4.3 RECRUTAMENTO E CARACTERIZAÇÃO CLÍNICA DA AMOSTRA POPULACIONAL

Um total de 36 pacientes portadores de DPIP foram incluídos no estudo, após diagnóstico realizado por neurologistas experientes, de acordo com os critérios de diagnóstico do Banco de Cérebros da Sociedade de Parkinson do Reino Unido (TABELA 1) (20).

A caracterização do grupo controle foi realizada por geriatra em um grupo de indivíduos com idade superior a 60 anos que foram submetidos a uma avaliação da condição geral de saúde, em conjunto com um levantamento de histórico, relatado pelo próprio paciente, de casos de doença neurodegenerativas na família. Esta avaliação levou ao recrutamento de 17 pessoas sem histórico de qualquer doença neurodegenerativa, individual ou familiar.

Todos os indivíduos recrutados são oriundos do estado do Paraná, sendo 17 controles e 28 pacientes residentes em Curitiba e região metropolitana, enquanto que os 7 pacientes restantes foram oriundos do interior do Paraná. Todos os voluntários assinaram Termo de Consentimento Livre e Esclarecido (TCLE) (APÊNDICE A), aprovado pelo comitê de Ética em pesquisa da PUCPR. O TCLE foi apresentado e totalmente explicado, em linguagem não técnica, a todos os indivíduos convidados a participar, imediatamente antes da coleta de amostra.

4.4 COLETA DE AMOSTRA E EXTRAÇÃO DE DNA

Os participantes foram submetidos a uma punção venosa para coleta de 10 mL de sangue periférico. As amostras de sangue foram coletadas em tubos com pressão reduzida contendo EDTA como anticoagulante, em condições ideais de assepsia.

A extração de DNA foi realizada por *salting out* a partir de sangue total, conforme descrito por John *et al.* (Anexo A) (180).

4.5 QUANTIFICAÇÃO E PADRONIZAÇÃO DE AMOSTRAS

A quantificação do material genômico foi realizada por espectrofotometria ultra-violeta em equipamento NanoDrop® 2000 (Catálogo nº ND-2000-US-CAN; *Thermo scientific*), utilizando-se 2µL de amostra e leituras nos comprimentos de onda de 260 nm e 280 nm. Amostras que apresentaram razão das leituras a 260/280 nm entre 1,5 e 2,0 foram consideradas de boa qualidade (109). Além deste parâmetro, foi utilizada a razão entre as absorbâncias de 260/230nm para a avaliação de qualidade quanto a presença de contaminantes. Uma razão de acima de 2 foi considerada adequada. Após a etapa de quantificação, todas as amostras foram padronizadas em soluções de trabalho com concentração de 20ng/µL. Tanto as soluções de trabalho como as amostras estoque foram armazenadas em freezer a -20°C, localizados em laboratórios da PUCPR, de acesso restrito aos investigadores deste estudo e ao pessoal técnico autorizado.

4.6 DESENHO DOS INICIADORES (*PRIMERS*)

As sequências utilizadas como referência neste trabalho foram obtidas da base de dados Emsembl (148-150), genoma GRCh37.p10 (181), e os *primers* foram desenhados com auxílio do programa *Primer3* (182-184) (APÊNDICE B). As sequências-alvo incluíram todas as sequências codificantes dos três genes estudados, sendo a sequência referência sempre a de maior extensão de DNA e proteína codificada, de forma a incluir todos os exons presentes nas diferentes isoformas. Os *primers* foram desenhados de forma a render amplicons com pelo menos 150 pares de bases flangeadoras (introns e as sequências 5' e 3' não traduzidas - *UnTranslated Regions, UTR*) da sequência-alvo. As condições de desenvolvimentos dos *primers* no programa foram as padrões exceto por: i) limite de extensão variável entre 18 e 22 pb; ii) temperatura de fusão (*Melting temperature, T_m*) entre 55 °C e 65 °C e iii) conteúdo CG próximo a 50%. O Apêndice C apresenta a sequência dos *primers* desenhados, os tamanhos dos amplicons e suas

localizações específicas no DNA.

Para testar a especificidade dos *primers* desenvolvidos, utilizou-se a ferramenta PCR *in silico* disponível no pacote de serviços do UCSC *genome Browser* (APÊNDICE B) (185). Todos os pares de *primers* apresentaram apenas um alvo teórico.

4.7 AMPLIFICAÇÃO DE DNA

Após a extração de DNA, todos os exons e segmentos das regiões 5' e 3' UTR dos genes *PARK2*, *PACRG* e *QKI* foram submetidas a amplificação por Reação em Cadeia da Polimerase (acrônimo em inglês de *Polymerase Chain Reaction*, PCR). Para cada indivíduo, a reação de PCR foi realizada para um volume final de 25µL, sendo deste volume 18µL água ultra pura, 2,5µL tampão 10X, 0,75µL de solução de cloreto de magnésio 50 mM, 0,75µL de *primer* direto 10pM, 0,75µL de *primer* reverso 10pM, 0,5µL de uma solução mix de dNTPs 10mM, 0,2µL de solução de *Taq* polimerase 5U/µL e 1,5 µL de uma solução de DNA 20 ng/µL.

As condições de reação programadas foram: 95°C por 3 minutos, seguido de 38 ciclos a 95°C por 30 segundos, temperatura específica para cada par de *primers* (APÊNDICE D) por 30 segundos, e 72°C por 60 segundos. Após os 38 ciclos era realizada uma etapa final de 72°C por 5 minutos e ao termino da reação seguia-se uma etapa de resfriamento até 4°C.

4.8 QUANTIFICAÇÃO, PADRONIZAÇÃO DE SOLUÇÃO DE AMPLICONS E POOLING

Uma alíquota de 2 µL dos produtos de PCR foram misturadas a 4 µL de uma solução de carregamento que contém corante azul de bromofenol, xileno cianol e *GelRed™* para então ser submetida a uma corrida eletroforética em gel de agarose a 2%, para controle de amplificação e quantificação em presença de marcador de massa molecular. A quantificação foi realizada por comparação da intensidade de

fluorescência entre as bandas do marcador de massa molecular e as bandas dos amplicons de massa desconhecida. Essa etapa foi realizada no sistema de captura de imagem e transiluminador Gerliance 200, com o auxílio dos *softwares* *GeneSnap™* para captura de imagem e *GeneTools™* (*Perkin Elmer*) para análise de fluorescência.

A padronização equimolar das soluções de amplicons foi realizada por diluição em água ultra pura, para uma concentração de 20 nM/μL. Cinco μL de cada solução de amplicons a 20 nM/μL de pacientes e controles foram combinadas em dois *pools*. Cada *pool* foi purificado por método de coluna *PureLink® PCR Purification Kit* (Catálogo nº K3100-01; *Life Technologies*) de acordo com as instruções do fabricante (número de publicação: MAN0000446; *Life Technologies*).

4.9 CONCATENAÇÃO DE AMPLICONS

Neste trabalho, a extensão dos amplicons variou de 181 pb a 805 pb. Assim, a fim de se gerar moléculas de DNA de tamanho uniforme, foi necessária a concatenação aleatória dos amplicons. Neste processo, uma ligase é utilizada para a produção de moléculas de DNA com comprimento variando entre 500 a 2000 pb que, em seguida, são fragmentadas em moléculas menores e de tamanho uniforme de aproximadamente 250 pb. A concatenação dos amplicons de cada *pool* (pacientes e controles) foi realizada de acordo com o protocolo *Applied Biosystems SOLiD™ System: Amplicon Concatenation Protocol* disponível na web ou mediante pedido ao suporte da *Life Technologies™*.

4.10 PREPARO DAS BIBLIOTECAS

O objetivo nesta etapa é ligar os amplicons a adaptadores universais que servirão para garantir a amplificação de todas as moléculas de maneira uniforme. Após etapa de ligação, é feita uma reação de reparo de quebra/fenda (*nick-repair*) para completar a ligação entre adaptadores e amplicons. A biblioteca preparada com

adaptadores é submetida a uma eletroforese para selecionar a extensão ótima de acordo com o tamanho da leitura (*read*) que se pretende gerar. De acordo com o protocolo, para a geração de *reads* de 200bp é necessária a seleção de fragmentos com extensão de 330pb em uma etapa de seleção de fragmentos por eletroforese. Estes fragmentos foram quantificados e submetidos a uma nova reação de PCR. Os amplicons gerados são novamente quantificados e utilizados na preparação de moldes.

As bibliotecas foram geradas usando o *Ion Xpress Plus fragment library kit* (Catálogo nº 4471269; *Life Technologies*) e *Ion Plus Fragment Library Adapters* (Catálogo nº 4476340; *Life Technologies*) conforme orientação do fabricante (186). Segue abaixo uma breve descrição das etapas necessárias para o preparo das bibliotecas:

4.10.1 Fragmentação do DNA

Esta etapa compreende a fragmentação química do DNA genômico em moléculas com tamanho médio de 330 pb que depois terão suas extremidades reparadas e fosforiladas para permitir a ligação de adaptadores universais de PCR. Após fragmentado e reparado, o DNA resultante é purificado utilizando método de micro-coluna. Cada *pool* foi submetido a este processo de maneira individual.

A fragmentação química do DNA foi realizada utilizando 100ng de cada *pool* de amplicons e 50µL de uma solução contendo 5µL de *Ion Shear Plus 10X reaction buffer*, 10µL *Ion Shear™ Plus Enzyme Mix II* e 35µL de água. Essa solução foi incubada a 37°C durante 20 minutos e na sequência a reação foi interrompida utilizando 5µL de *Ion Shear stop buffer*. O DNA fragmentado foi purificado usando *Agencourt Ampure XP-PCR purification beads* (PN A63880; Beckman Coulter, Brea, CA) com *Dynal magnetic bead stand* (Catálogo nº123-21D; *Life Technologies*). Resumidamente, 99µL de microesferas *Agencourt* foram misturadas com 50µL do DNA fragmentado e a mistura foi incubada por 5 minutos a temperatura ambiente. Após a incubação, colocou-se a solução em um suporte magnético, lavou-se duas vezes com etanol 70% e finalmente eluiu-se utilizando 25µL de *low-TE buffer* (Catálogo nº 602-1155-010; *Life Technologies*).

4.10.2 Ligação dos adaptadores e *nick-repair*

Para realizar a ligação dos adaptadores, misturou-se 25µL do DNA fragmentado, 10µL de tampão ligase, 2µL de adaptadores (*Ion Plus Fragment Library Adapters*), 2µL de DNA ligase (*Ion Xpress Plus fragment library kit*), 2µL de mix de dNTPs, 8µL de *nick repair polymerase* (*Ion Xpress Plus fragment library kit*) e 51µL de água livre de nuclease. Incubou-se a mistura em termociclador a 25°C por 15 minutos, a 72°C por 5 minutos e 4°C até a retirada a amostra do equipamento. Finalmente, purificou-se utilizando a mesma metodologia de purificação do DNA fragmentado, conforme descrito na etapa anterior, utilizando microesferas *Agencourt* e suporte magnético. Após a purificação, eluiu-se a amostra utilizando 20µL de *low-TE buffer*.

4.10.3 Seleção de tamanho dos fragmentos

Para a seleção de moléculas de DNA de tamanho específico para o sequenciamento, realizou-se uma eletroforese em gel de agarose a 2% *E-Gel® SizeSelect™* (Catálogo nº G6610-02; *Life Technologies*). Este gel apresenta duas linhas de poços: na superior, adiciona-se a amostra para a corrida eletroforética; na segunda linha de poços – chamados de poços coletores – recuperam-se as moléculas de DNA com o tamanho de interesse. Para isso, adicionou-se toda a amostra de fragmentos de DNA ligados a adaptadores nos poços superiores, e para estimar o tamanho das bandas utilizou-se um marcador de peso molecular. Para a corrida eletroforética, utilizou-se o programa *SizeSelect™ 2%* no *E-Gel® Power Base* e acompanhou-se a corrida pelo *E-Gel® Safe Imager™ Real-Time Transilluminator* (*E-Gel® iBase™ and E-Gel® Safe Imager™ Combo Kit*, catálogo nºG6465; *Life Technologies*). Para a recuperação de fragmentos de DNA de 350 pb, interrompeu-se a corrida no momento em que a banda de 350 pb do marcador de peso molecular se encontrava na porção superior do poço coletor, e retirou-se com uma pipeta as soluções dos poços coletores correspondentes às amostras.

4.10.4 Amplificação e purificação das bibliotecas

Esta etapa consiste de uma PCR comum na qual todos os fragmentos que possuem os adaptadores universais ligados vão ser amplificados. Para isso, combinou-se em um tubo 200µL de *Platinum® PCR Super Mix High Fidelity*, 10µL de *Library Amplification Primer Mix* e 50µL da biblioteca. Colocou-se esta solução em um termociclador e iniciou-se o ciclo de PCR com os seguintes parâmetros: estágio inicial de desnaturação a 95°C por 5 minutos, 5 ciclos de desnaturação a 95°C por 15 segundos seguido de anelamento a 58°C por 15 segundos e extensão a 70°C por 1 minuto, e estágio final a 4°C por tempo indeterminado. Após a PCR realizou-se a purificação da biblioteca semelhante às realizadas nas etapas anteriores.

4.10.5 Quantificação e diluição das bibliotecas amplificadas

Após a amplificação das bibliotecas, é necessário quantificar o produto da PCR para determinar o fator de diluição que resulte na concentração ideal para a preparação dos moldes. A quantificação foi realizada por corrida microfluídica na plataforma BioAnalyzer 2100 (*Agilent Technologies*) com a utilização do *Agilent® High Sensitivity DNA Kit* (Catálogo nº 5067-4626; *Agilent Technologies*). A concentração foi determinada em pmol/L usando o *Bioanalyzer® software*, e a partir dela foi calculada a diluição necessária para se obter uma concentração de aproximadamente 26 pM da biblioteca.

4.11 PREPARO DOS MOLDES

O objetivo desta etapa é preparar as moléculas da biblioteca de DNA para

serem sequenciadas, através da fixação das mesmas ao suporte de reação. Para isso, a biblioteca foi amplificada novamente, mas desta vez em um meio emulsificado. É preparada uma solução aquosa com a biblioteca, polimerase e deoxinucleotídeos na presença de microesferas que apresentam um *primer* universal ligado à sua superfície. A biblioteca, por sua vez, tem em suas extremidades sequências que hibridam com o *primer* da microesfera. Desta forma, durante a reação, novas cadeias de DNA fita simples serão amplificadas, gerando alvos fisicamente ligados a microesfera. Para que isso ocorra, o meio aquoso preparado deve ser emulsionado em uma solução oleosa de modo que a emulsão gere uma gotícula de água contendo uma única microesfera, apenas um fragmento de DNA e os demais componentes de uma *PCR* clássica. Nestas condições tem-se milhares de microreatores que, idealmente, deverão amplificar apenas um tipo de amplicom em cada sistema, gerando uma *ISP* clonal. Depois da amplificação, a emulsão é quebrada, as *ISP* carregadas são lavadas e separadas das *ISP* vazias (enriquecimento). Desta forma, tem-se moldes prontos para a preparação e carregamento do chip de sequenciamento.

O preparo dos moldes (moléculas que são efetivamente sequenciadas) foi realizado usando o *Ion Xpress™ Template Kit* (Catálogo nº 4469001; *Life Technologies*) conforme orientação do fabricante (manual 4469004 Rev. E). O preparo dos moldes envolve as seguintes etapas apresentadas a seguir:

4.11.1 PCR em emulsão (ePCR)

Na ePCR, são preparadas duas fases, uma fase aquosa contém as moléculas-alvo, microesferas, *primers*, enzimas e demais reagentes necessários para a reação de PCR e a outra uma fase oleosa que serve para aprisionar a solução aquosa durante a emulsão e reação de PCR, formando microreatores. Cada microrreator funcionará como um, produz muitas cópias idênticas de um único fragmento ligado na superfície da microesfera. Resumidamente, em um volume final de 1mL, adicionou-se 582µL de água livre de nuclease, 200µL de *5X PCR reagent mix*, 100µL de *10X PCR enzyme mix*, 100µL de *Ion Sphere particles* (ISPs), e 18µL de *diluted library template*. Misturou-se vigorosamente a solução, que em seguida foi

submetida a uma breve centrifugação. Na sequência, colocou-se 9mL de óleo de emulsão refrigerado (*Ion template preparation kit*; *Life Technologies*) em um tubo denominado *Ion template preparation tube* (Catálogo nº 4467226; *Life Technologies*), específico para o preparo da emulsão pelo equipamento *Ultra-Turraxtube drive* (*Life Technologies*). Para isso, colocou-se o tubo com o óleo no equipamento e, enquanto a solução estava em movimentação, adicionou-se lentamente a solução aquosa pelo orifício da tampa do tubo e misturou-se durante 5 minutos. Depois, a emulsão foi transferida para uma placa de PCR de 96 poços e amplificada no termociclador usando os seguintes parâmetros: 94°C por 6 minutos, 40 ciclos a 94°C por 30 segundos, 58°C por 30 segundos, e 72°C por 90 segundos; finalmente, 5 ciclos a 94°C por 30 segundos e 68°C por 6 minutos.

4.11.2 Quebra da emulsão e lavagem das microesferas

As microesferas foram recuperadas utilizando o *Ion Xpress template kit* (Catálogo nº 4469001; *Life Technologies*); na sequência, realizou-se a quantificação das microesferas recuperadas usando fluorômetro Qubit 2.0 (*Life Technologies*) de acordo com as recomendações do fabricante.

4.11.3 Enriquecimento das microesferas carregadas

Este procedimento isola as microesferas carregadas das vazias por hibridação de oligonucleotídeos, presentes nas microesferas de enriquecimento, com as cadeias de DNA presas nas microesferas carregadas. Desta forma, são retidas de 150 à 300 milhões de microesferas carregadas que serão utilizadas na etapa seguinte.

Para tal, utilizou-se o *Ion Xpress template kit* (Catálogo nº 4469001; *Life Technologies*), *Ion sequencing kit* (Catálogo nº 4468997; *Life Technologies*), e *DynaBeads MyOne streptavidin C1 beads* (Catálogo nº 650.01; *Life Technologies*).

4.12 SEQUENCIAMENTO

Nesta etapa, uma suspensão de microesferas carregadas com amostras é adicionada de *primers* e polimerase de sequenciamento para então serem injetadas no chip de sequenciamento. Carregado, o chip é inserido na plataforma e um programa de sequenciamento é definido.

Para realizar o sequenciamento dos *pools* na plataforma *Ion PGM*, utilizou-se o *Ion PGM Sequencing 200 Kit v2* (Catálogo nº 4482006) com o chip *Ion 314*, seguindo as indicações do fabricante (manual MAN0007273, *Life Technologies*).

4.13 ANÁLISE DOS DADOS

Os dados brutos de sequenciamento produzidos pela plataforma *Ion PGM* foram analisados em um processo que pode ser dividido em três etapas – chamadas de análise primária, secundária e terciária (128, 187). Os componentes de cada etapa estão apresentados na forma de fluxograma na figura 4, e os procedimentos são apresentados em detalhes nos tópicos a seguir. Novamente, para fins didáticos, os objetivos de cada fase das três etapas são descritos brevemente.

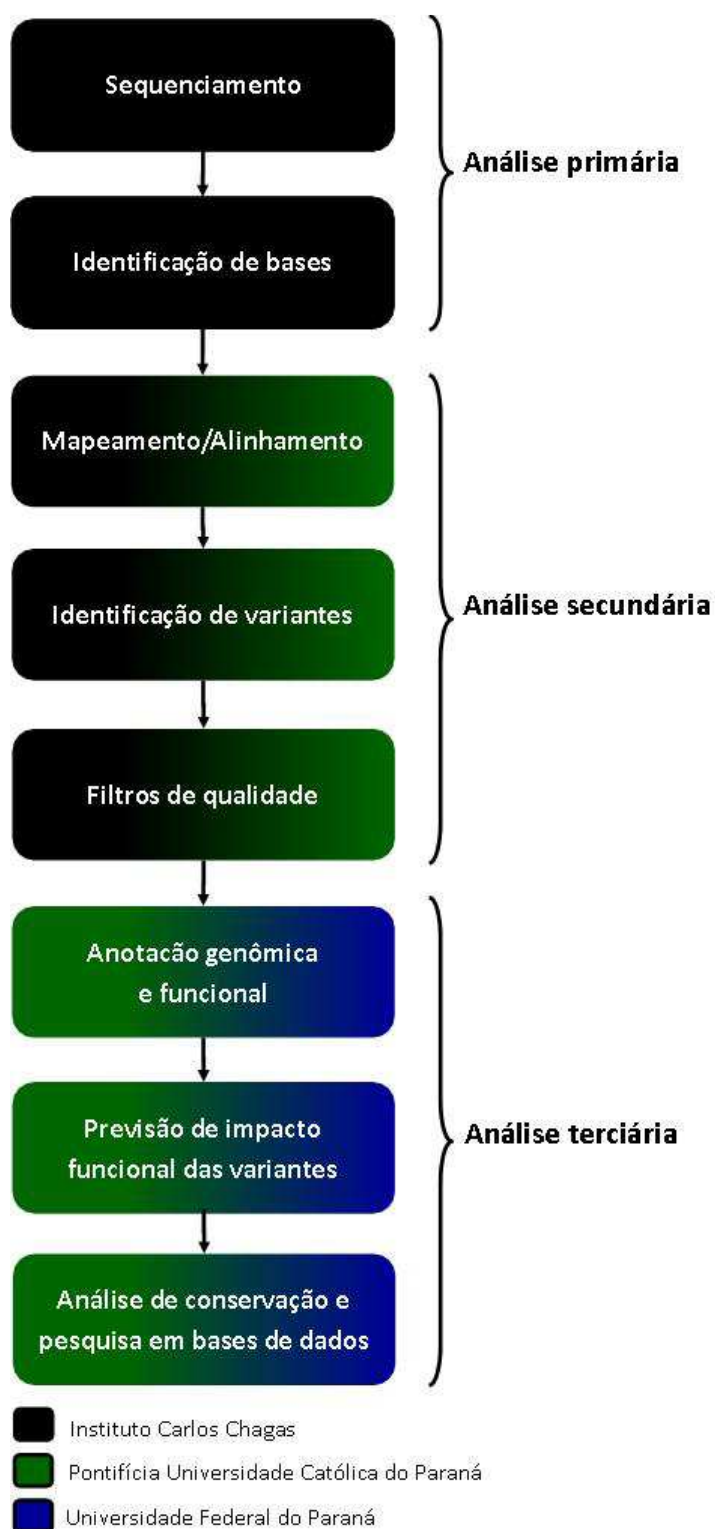


FIGURA 4. Esquema da análise de dados. As cores indicam a(s) instituição(ões) envolvida(s) em cada etapa.

4.13.1 Análise primária

Esta etapa tem a finalidade de transformar os sinais brutos obtidos durante o sequenciamento nas sequências curtas, ou *reads*. As *reads* são obtidas através de um conjunto de algoritmos aplicados em sequência (*analysis pipeline*) que interpretam o sinal elétrico produzido pela mudança de pH do sistema de reação (liberação de H^+), e os transformam em um arquivo que contém, além das *reads*, informações (*score*) de qualidade para cada nucleotídeo incorporado durante a amplificação dos alvos em cada poço. Os dados de qualidade informados são baseados no *score Phred*, a medida padrão para determinar a qualidade da informação de cada posição de nucleotídeo identificado nas sequências de DNA (188, 189), e dados de *Flow order*. O arquivo de saída produzido após a análise primária é gerado no formato BAM, que armazena todas as *reads* independentemente de informações de um genoma de referência.

Os algoritmos utilizados nesta etapa foram os fornecidos para a plataforma *Ion PGM* (Torrent Suit) e foram empregados utilizando configurações padrão. Detalhes sobre quais são e o modo como operam estes algoritmos podem ser encontrados na sessão *Technical Note - Analysis Pipeline* do documento *TSUD_3.4.1_User Guide*. Todo o processamento das sequências e geração de arquivo final não mapeado BAM são realizados sem a interferência do operador.

4.13.2 Análise secundária

A análise secundária tem como objetivo reconstruir a sequência original a partir da qual as *reads* foram geradas, um processo conhecido como alinhamento ou mapeamento. Nesta etapa, o processamento das informações do arquivo BAM é feito com o auxílio do programa *The Torrent Mapping Alignment Program for Ion Torrent™* (TMAP). Foram utilizadas as configurações padrões para o alinhamento; no entanto, a sequência referência foi adaptada a partir do genoma de referência

GRCh37.p10 de agosto de 2012 (181). Deste *assembly* foram extraídas sequências completas do genes *PARK2*, *PACRG* e *QKI*, a partir das quais foi montada uma sequência referência contendo apenas as regiões codificantes dos exons e suas sequências flanqueadoras (introns e/ou UTRs), compatíveis como os alvos esperados gerados a partir das reações de *PCR*. As regiões alvo foram separadas por espaçadores "N" de 100 pb. Esta medida foi tomada para se evitar uma demanda computacional excessiva, uma vez que não são esperados alinhamentos fora das regiões alvo.

Para a identificação de variantes, foram empregadas as ferramentas *Torrent variant caller* (TVC), parte do conjunto de *softwares* da plataforma, e o *Integrative Genome Viewer* (IGV) versão 2.2.7 (190, 191). O algoritmo TVC é utilizado na identificação de variantes. Foram empregados os parâmetros padrões do protocolo de identificação de variantes somáticas, exceto pela opção *het-min-allele-ratio*, que foi ajustado de 4% para 1%. Este parâmetro refere-se à frequência mínima observada nas *reads* a partir da qual uma alteração de bases pode ser identificada como variante. O recomendado pelos desenvolvedores é uma frequência de no mínimo 3%, uma vez que abaixo deste valor a chance de se identificar variantes falsas é inflacionada. No entanto, optou-se pela frequência de 1% em virtude da abordagem de *pooling*, pois dessa forma uma variante presente em apenas um cromossomo de um indivíduo no *pool* paciente contribui com uma frequência de apenas 1,38% na ocorrência de um SNV enquanto que para o grupo controle a frequência mínima esperada era de 2,94%.

Sabendo-se que: i) as plataformas de sequenciamento NGS são propensas a uma maior taxa de erro de sequenciamento quando comparadas com as plataformas clássicas baseadas no método de Sanger (175), e ii) a alteração no limiar de detecção de SNV pode inflacionar a identificação de falsos positivos, todas as variantes identificadas neste estudo foram consideradas candidatas que precisarão ser validadas por outra plataforma de sequenciamento. No entanto, antes de serem realizados experimento de validação, faz-se necessário a aplicação de métodos de triagem de variantes com alta probabilidade de serem reais, com consequente diminuição de falsos positivos.

Desta forma, faz-se necessário uma análise conjunta com o programa IGV, uma ferramenta de visualização dos alinhamentos gerados, que permite

modificações em seu sistema de filtros para auxiliar na tomada de decisão sobre falsos positivos e/ou variantes mais prováveis para a posição variável identificada. Os parâmetros utilizados para esta análise foram os padrões do programa, com exceção do filtro de qualidade de mapeamento. As *reads* foram filtradas de acordo com uma qualidade de mapeamento na escala *Phred* no valor de 20, ou seja, 99% de probabilidade de um mapeamento correto. Partindo-se do alinhamento nestas condições e dos parâmetros de qualidade que o programa apresenta para cada variante, foram aplicados os seguintes critérios para a inclusão de variantes como potencialmente verdadeiras:

- **Razão entre as sequências diretas e reversas (Dir:Rev):** Este parâmetro é uma razão entre a contagem do número de *reads* não clonais nos sentidos direto e reverso. Estas razões tiveram como numerador a contagem mais alta de cobertura da variante candidata. Se uma variante for verdadeira, espera-se que ela seja igualmente representada em ambos os sentidos de leitura. Neste trabalho, o corte máximo adotado foi de 2,0 para considerar uma variante candidata viável e não um artefato.
- **Qualidade de mapeamento:** Este indicador refere-se à qualidade de mapeamento de cada *read* que continha uma variante identificada e é apresentado pelo programa TVC, como uma média da qualidade de todas as *reads*. Para o cálculo deste parâmetro, utiliza-se o número de bases que diferem da referência, a escala *Phred* destas e a probabilidade de a *read* em questão encontrar alinhamento em outras áreas do genoma (192). Como inicialmente adotou-se o corte de qualidade de mapeamento de Q20, variantes com valores inferiores a este foram excluídas da análise.
- **Qualidade de identificação de Variantes (Q_Var):** Este parâmetro é baseado em uma escala *Phred-like* que reúne informações referentes à qualidade de mapeamento da *read* que contém uma variante (descrito acima) e a qualidade *Phred* e cobertura da variante, para estimar a probabilidade dela ser um artefato. Este parâmetro é calculado conforme implementado pelo

programa TVC. Foi adotado o valor de corte de Q_Var 30, uma vez que esse valor representa chance de erro de 0,1% na identificação da variante.

- **Cobertura:** Tendo em vista que a cobertura média das sequências foi de 3.000X, optou-se por uma linha de corte segura, de cobertura mínima de 600X. Este corte foi adotado devido à observação de que regiões com cobertura inferior a esta, 90% das vezes apresentavam-se em áreas de cobertura comprometida. Estas são áreas que cobrem o final das sequências referência e que apresentam falhas (*gaps*) de coberta (FIGURA 5B).
- **Frequência do alelo alternativo nas leituras:** A variante candidata tem maior probabilidade de ser verdadeira quanto mais alta for sua frequência nas *reads*. No entanto, essa medida não foi adotada como valor de corte, mas sim, como dado de auxílio na tomada de decisão.

As variantes que passaram pelos critérios citados acima foram também avaliadas quanto à presença de artefatos (exemplo, FIGURA 2) e áreas de homopolímeros nas suas *reads* de origem: além da variante identificada que passou o filtro de qualidade, estas *reads* possuíam mais de 3 bases diferentes da referência em toda a extensão da *read* (com escala *Phred* inferior a 30) e áreas de homopolímeros – áreas de pouca resolução para a plataforma *Ion PGM* (193).

Os parâmetros acima sempre foram utilizados em conjunto, e variantes candidatas foram consideradas verdadeiras quando apresentavam o seguinte perfil:

- Razão Dir:Rev abaixo de 2,0,
- Q_Var acima de 30,
- Cobertura superior a 600X, e
- Ausência de artefatos e área de homopolímeros nas mesmas *reads* que a variante.

Exceções foram consideradas somente nas seguintes situações:

- Variantes com razão Dir:Rev entre 2,0 e 3,0 foram consideradas verdadeiras somente se apresentavam Q_Var acima de 40, cobertura superior a 1.000X, frequência do alelo alternativo nas *reads* superior a 2,0% e ausência de artefatos e área de homopolímeros nas mesmas *reads* que a variante.
- Variantes com cobertura entre 100X e 600X foram consideradas verdadeiras somente se apresentavam razão Dir:Rev inferior a 2,0, Q_Var acima de 40, frequência do alelo alternativo nas *reads* superior a 2,0% e ausência de artefatos e área de homopolímeros nas mesmas *reads* que a variante.

A figura 5A ilustra um alinhamento de alta cobertura (3.520X), destacando um SNV identificado pelo TVC, que foi validado pelos critérios estabelecidos. Observa-se que o alelo referência é o T e o alternativo, C. A figura 5B ilustra uma área de leitura comprometida (a posição em destaque tem 538X), com vários *gaps*, artefatos e uma variante identificada pelo TVC que não foi considerada verdadeira pelos critérios de inspeção.

Finalmente, todas as variantes consideradas verdadeiras de acordo com o protocolo descrito acima passaram por um processo de anotação manual (*database annotation*), e foram distribuídas em duas classes: variantes novas e variantes previamente descritas, presentes nas bases de dados *NCBI dbSNP* (194), *1000 Genomes* (195, 196) e *HapMap* (151, 152) (APÊNDICE B).

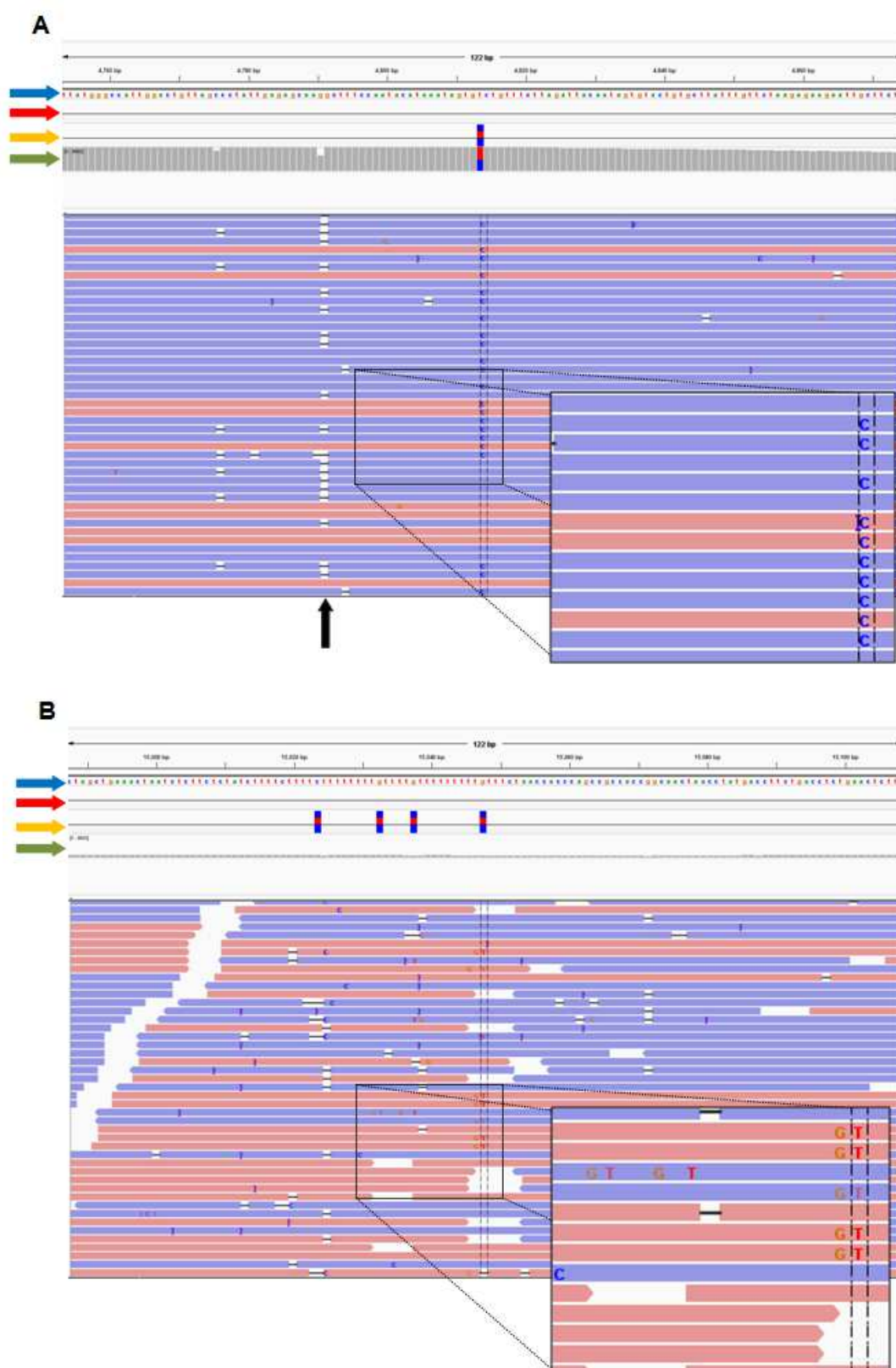


FIGURA 5. Visão esquemática do alinhamento e cobertura de duas regiões com qualidades diferentes segundo o algoritmo IGV. **A)** Exemplo de uma região com perfil adequado **B)** Exemplo de área com cobertura comprometida e artefatos. Seta azul: Sequência referência. Seta vermelha: Campo sinalizador de indels. Seta amarela: Campo sinalizador de variações de ponto. Seta verde: Campo sinalizador de densidade de cobertura. Seta preta: Erro de sequenciamento - deleção. Em rosa destacam-se as *reads* diretas e em azul as *reads* reversas.

4.13.3 Análise terciária

Nesta etapa, são atribuídas às variantes informações acerca de sua localização, categoria e impacto funcional. A previsão de impacto funcional de variantes na estrutura da proteína foi realizada com os algoritmos *Polymorphism phenotyping 2* (PolyPhen-2) (197, 198) e *Sorting intolerant from tolerant* (SIFT) (199, 200).

PolyPhen-2 prediz o efeito de uma substituição de aminoácido sobre a estrutura e função da proteína utilizando homologia de sequências, anotações de famílias de proteínas (Pfam), entre outras bases de dados. Utilizando estes dados o programa gera um *score* e atribui a ele uma previsão qualitativa que pode ser classificada em "provavelmente deletério", "possivelmente deletério", "benigno" ou "desconhecido". O *score* PolyPhen-2 é apresentado em uma escala que vai de 0 a 1; a probabilidade de um efeito deletério é tanto maior quando mais próximo o valor chegar de 1.

SIFT tem a mesma função de PolyPhen-2, no entanto, utiliza apenas homologia de sequências e a similaridade entre as propriedades físico-químicas dos aminoácidos substituídos para gerar previsões. A previsão qualitativa ("tolerável" ou "deletéria") é baseada em um *score* que também varia de 0 a 1. No entanto, a interpretação é oposta à aplicada na escala PolyPhen-2: valores SIFT abaixo de 0.05 indicam substituição deletéria, e quanto mais o valor se aproximar de zero, mais deletéria é a substituição.

Além dessas análises, foram realizadas investigações de possíveis sítios de *splicing* pelos programas *Human Splice Finder* (HSF) versão 2.4.1 (201) e *Alternative Splice Site Predictor* (ASSP) (202), e de possíveis sítios de regulação da expressão em RNAm (sítio de ligação de miRNA ou alteração de regiões importantes em UTRs) com o auxílio do programa RegRNA 2.0 (203) (APÊNDICE B).

Além disso, avaliou-se a conservação das sequências de aminoácidos entre as espécies *Homo sapiens* (humano), *Pan troglodytes* (chimpanzé), *Mus musculus* (camundongo), *Rattus norvegicus* (rato), *Gallus gallus* (galinha), *Danio rerio* (peixe-

zebra) e *Drosophila melanogaster* (mosca-da-fruta), utilizando sequências referências extraídas da base de dados NCBI (148-150). O alinhamento destas sequências foi gerado pelos programas *Multiple Sequence Comparison by Log-Expectation* (MUSCLE) (204, 205) e Clustal Omega (206). Todos os programas citados são de livre acesso na web e foram utilizados em suas configurações padrão.

Finalmente, foi realizado levantamento bibliográfico buscando relatos anteriores sobre o envolvimento das variantes detectadas em DP e DPIIP. Além das bases de dados NCBI, Ensembl e HapMap, buscas adicionais foram realizadas nas bases *Human Gene Mutation Database* (HGMD) (207, 208) e *Leiden Open Variation Database* (LOVD) (209) (APÊNDICE B).

5 RESULTADOS E DISCUSSÃO

5.1 DESCRIÇÃO DA POPULAÇÃO

O grupo de pacientes portadores de DPIP apresentou idade média de recrutamento e diagnóstico de 49,6 e 32,5 anos, respectivamente (TABELA 3). A idade média do grupo controle foi de 66,4 anos (TABELA 3), que caracteriza o grupo como sendo livre de DPIP, embora não como livres de desenvolver DP idiopática, que se manifesta tipicamente entre os 60 e 70 anos (2, 8). Houve um discreto aumento na proporção de homens caucasianos no grupo de casos de DPIP, em comparação com o grupo controle (TABELA 3).

TABELA 3. Características demográficas e clínicas da amostra populacional.

Variável		Amostra populacional	
		Casos (n=36)	Controles (n=17)
Gênero	Masculino	29 (80,56%)	11 (64,70%)
	Feminino	7 (19,44%)	6 (35,30%)
Etnia	Branco	32 (88,89%)	13 (76,46%)
	Negro	1 (2,78%)	2 (11,77%)
	Pardo	3 (8,33%)	2 (11,77%)
Idade*	Média	49,6	66,4
	Mínima	30	62
	Máxima	77	82
Idade de diagnóstico*	Média	32,5	-
	Mínima	15	-
	Máxima	45	-

* Idade em anos.

5.2 ANÁLISE DOS DADOS

5.2.1 Análise primária

O sequenciamento dos *pools* pacientes e controles foi realizado em

experimentos independentes, o que gerou um perfil diferente de resultados em termos de rendimento e qualidade dos dados. Em geral, os dados obtidos do experimento de sequenciamento do *pool* controle foram de melhor qualidade quando comparados com o do *pool* pacientes. Esta ocorrência deve-se em parte às condições de carregamento do chip com as microesferas contendo em sua superfície as moléculas a serem sequenciadas (FIGURA 6). O carregamento do chip com alvos do *pool* controle (FIGURA 6B) foi mais eficiente em comparação ao *pool* paciente (FIGURA 6A). Esta condição foi provocada pela presença de uma bolha de ar dentro do chip de sequenciamento. Para o *pool* controle, o total de bases geradas a partir de um carregamento de 59% da área total foi de 65.9 Mb; para o *pool* pacientes, um carregamento de 41% da área total do chip gerou 32.2 Mb (FIGURA 6).

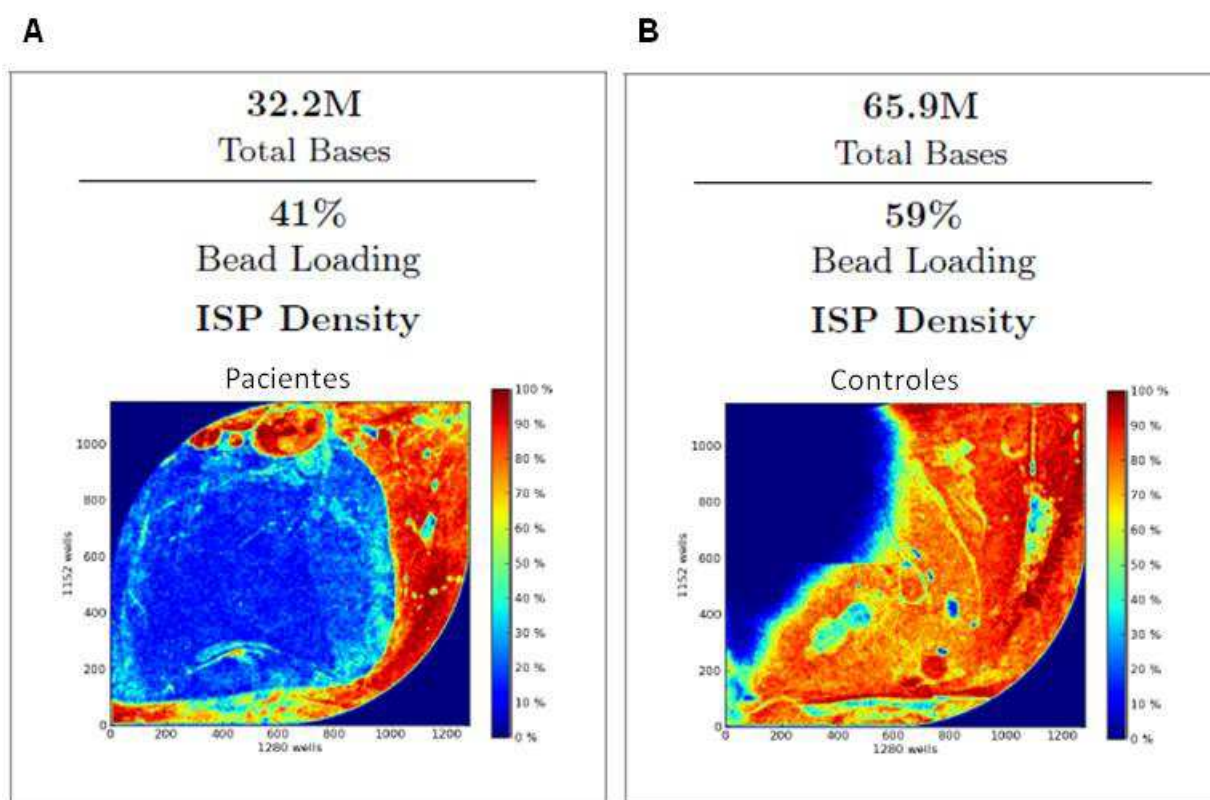


FIGURA 6. Imagens das áreas de carregamento contendo as microesferas (*Ion Sphere Particles*-ISP) com os *pools* de amplicons. **A)** Carregamento do *pool* paciente e **B)** do *pool* controle. A variação da escala de cores indica a densidade de poços carregados com ISP enriquecidas; cores tendendo para o vermelho indicam boa densidade, e cores tendendo para o azul, indicam baixa densidade de microesferas carregadas. M: Megabases

A figura 7 apresenta um resumo das condições de carregamento do chip. No *pool* casos, o total de microesferas carregadas foi de 518.677, que ocuparam uma

área de 41% do chip (FIGURA 7A.I). Destas, 81% continham moléculas de DNA (FIGURA 7A.II) e destas, 69% (290.129microesferas) eram clonais (FIGURA 7A.III), ou seja, microesferas com apenas um tipo de molécula de DNA que foi sequenciada em um poço específico. Das 290.129 microesferas clonais, 70% (204.330) geraram leituras de *reads* que foram utilizadas nas análises; os 30% excluídos correspondem a 26% retirados por serem de baixa qualidade e 4% continham sequências de teste de fragmento, sequências conhecidas e introduzidas junto com a biblioteca para análise de qualidade (FIGURA 7A.IV). Do total de bases geradas (32.2 Mb para pacientes) 54% geraram sequências adequadas (FIGURA 7A). Para o *pool* controle, a quantidade total de *reads* e, por consequência, de bases, foi superior ao *pool* paciente, sendo o número final de *reads* 369.108 (FIGURA 7B.IV), que corresponde a 54% do total de 65.9 Mb de bases geradas (FIGURA 7B).

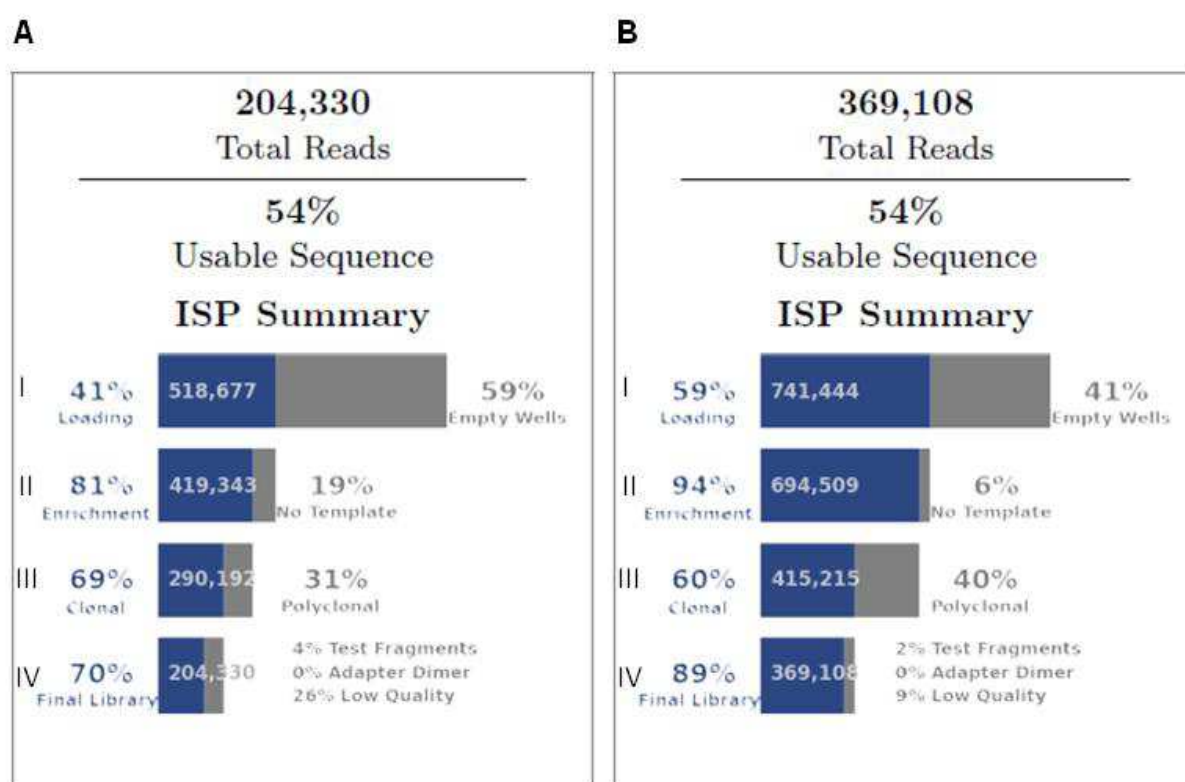


FIGURA 7. Rendimento das microesferas (*Ion Sphere Particles*- ISP) geradas. Resultados para o *pool* **A)** pacientes e **B)** controles. I) carregamento do chip, II) presença de microesferas enriquecidas, III) presença de microesferas clonais, IV) biblioteca efetiva.

A figura 8 apresenta o perfil de distribuição do tamanho das *reads* e seu número total por tamanho. Observa-se um perfil bi-modal de distribuição dos tamanhos das *reads* nos dois *pools*. Para o *pool* pacientes, parte das *reads* teve comprimento entre 70 e 220 pb, e parte, na faixa de 220 a 280 pb (FIGURA 8A). As

reads do *pool* controle também tiveram distribuição bi-modal, com um conjunto na faixa de 120 a 220 pb e outro com 240 a 280 pb (FIGURA 8B). Estes perfis refletem uma seleção não-uniforme na etapa final do preparo da biblioteca, uma vez que o perfil esperado é de um pico único na faixa entre 170 e 220 pb para um melhor rendimento e montagem inequívoca dos alinhamentos. Esta limitação técnica não comprometeu as análises posteriores, uma vez que este estudo teve como alvo apenas regiões específicas de DNA amplificadas por PCR, excluindo-se assim a possibilidade de geração de *reads* a partir de regiões homólogas; além disso, a limitação foi compensada pela alta cobertura obtida, de 3.000X em média.

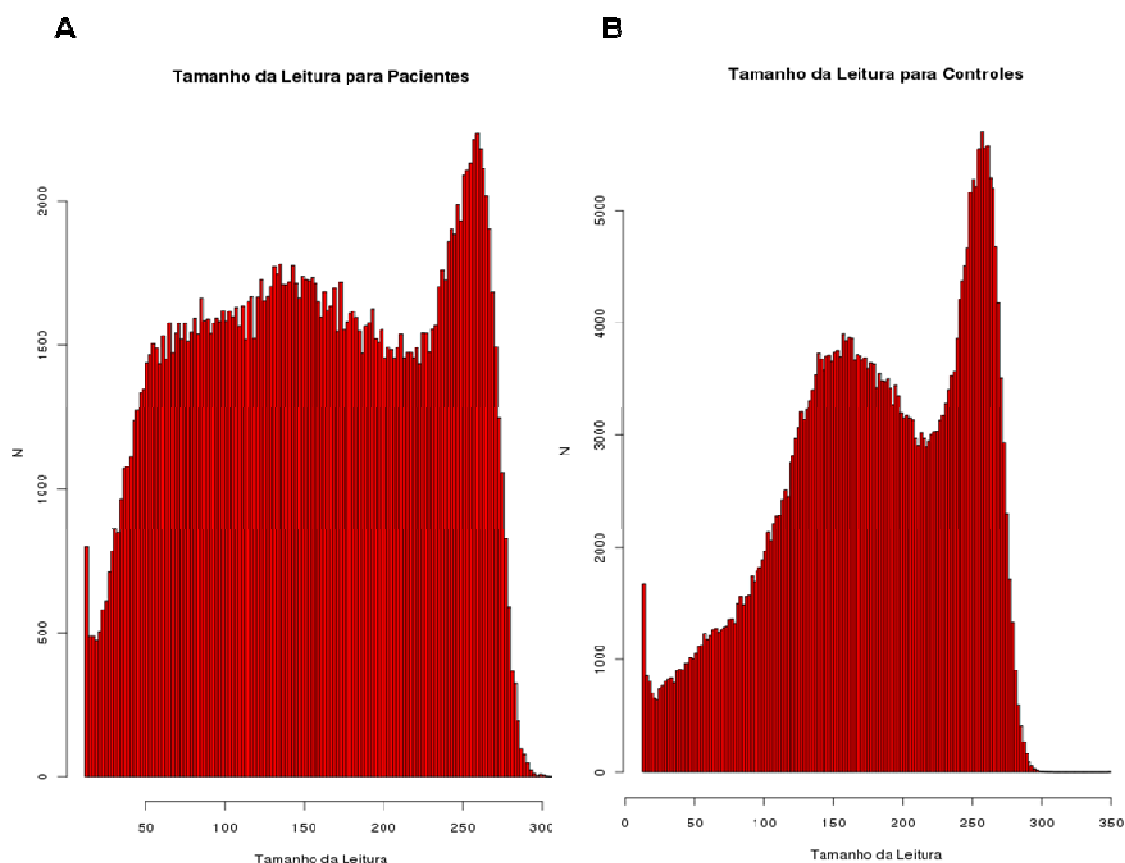


FIGURA 8. Distribuição dos tamanhos das leituras geradas. **A)** *Pool* pacientes e **B)** *pool* controle. O eixo y corresponde ao número de *reads* e o eixo x, o tamanho das *reads*.

5.2.2 Análise secundária

Um sumário dos parâmetros dos mapeamentos realizados com o programa TMAP e as estatísticas gerais são apresentadas na figura 9. A porcentagem de

reads mapeadas foi semelhante em ambos os *pools* (FIGURA 9.I), com acurácia superior a 97% em ambos os *pools* (FIGURA 9.II). A aparentemente alta proporção de bases não alinhadas (19% para o *pool* pacientes e 20% para o *pool* controles) é consequência do método de uniformização dos tamanhos dos amplicons: a concatenação e posterior fragmentação favorece a formação de *reads* quiméricas – ou seja, *reads* que combinam sequências de regiões exônicas diferentes – que previsivelmente não alinham com a sequência referência. Este artefato é esperado e não altera a qualidade final do alinhamento.

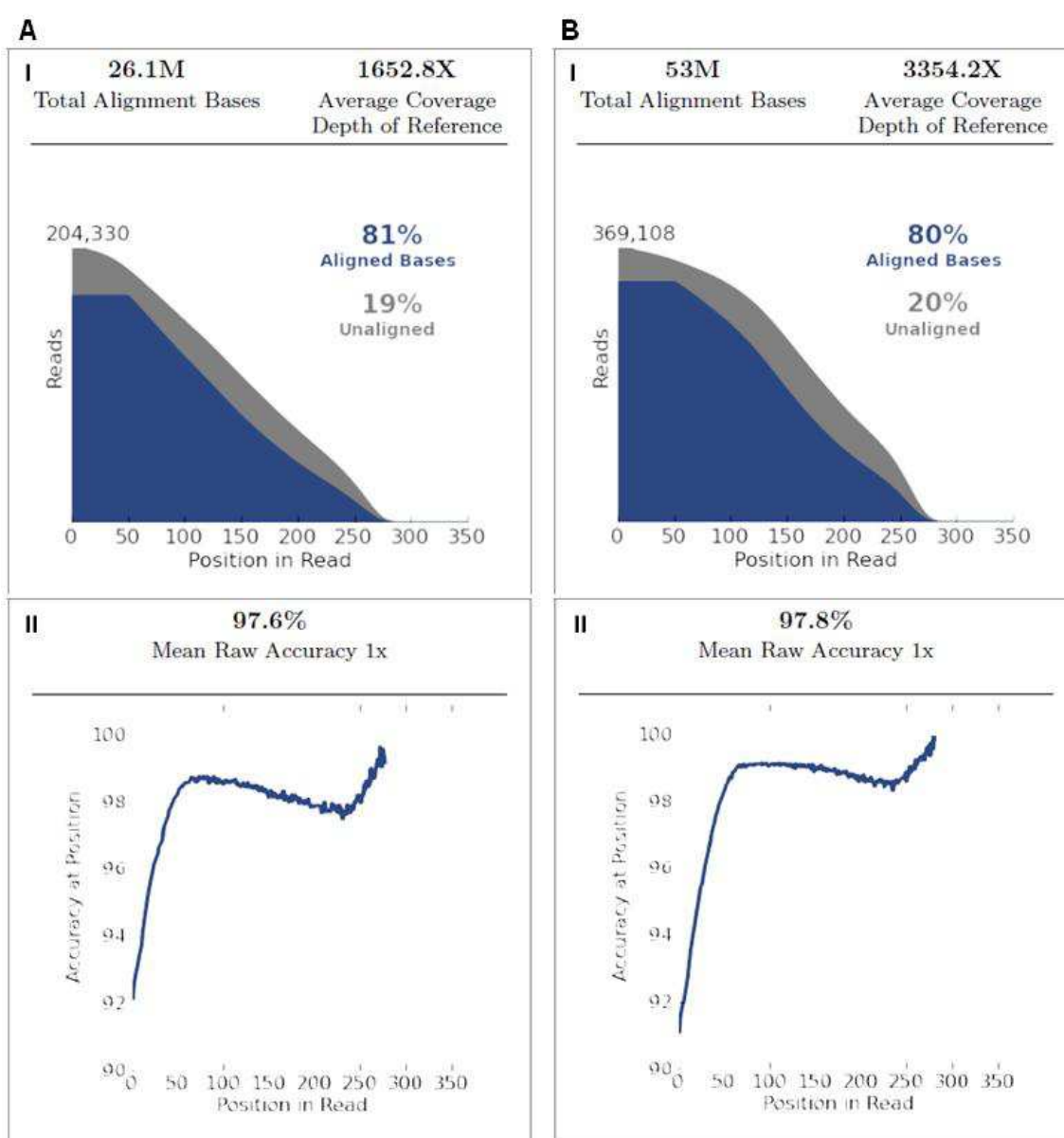


FIGURA 9. Estatística do total de *reads* mapeadas. **A)** *Pool* pacientes e **B)** *pool* controles. **I)** Porcentagens de bases alinhadas (azul) e não alinhadas (cinza), de acordo com a sua posição na *read*. **II)** porcentagem da acurácia de alinhamento das bases por categoria de *read*.

A tabela 4 apresenta os resultados de qualidade geral das *reads* e bases alinhadas. O parâmetro AQ (*Alignment Quality*) refere-se à qualidade de alinhamento em escala *Phred*; AQ17 corresponde a escala *Phred* de 17, ou seja, 2% de erro na identificação de bases em uma determinada *read* (isto é, 2 nucleotídeos incorretos em uma *read* de 100 pb); AQ20 equivale a 1% de erro e "perfeito" equivale a qualidade igual ou superior a AQ30 ($\leq 0,1\%$ de erro). Em geral, o resultado de alinhamento foi melhor para o *pool* controle, que apresentou 30,5 Mb de sequências com qualidade superior a AQ30; estas sequências correspondem a *reads* com comprimento médio de 113 pb sem nenhum erro de sequenciamento. O alinhamento mais extenso com qualidade AQ30 foi de 292 pb, e a cobertura média em uma base corretamente identificada foi de 1932,8X (número de *reads* que apresentam uma base com qualidade *Phred* 30 mapeando com a referência).

TABELA 4. Qualidade de alinhamento das *reads*.

	Controles			Pacientes		
	AQ17	AQ20	Perfeito	AQ17	AQ20	Perfeito
Número total de bases (Mb)	44,5	39,8	30,5	21,3	18	13,5
Comprimento médio (pb)	151	142	113	133	118	92
Alinhamento mais extenso (pb)	299	299	292	293	293	292
Cobertura média	2819,7	2525,1	1932,8	1353,1	1139,4	858,7

AQ17: *alignment quality* 17; AQ20: *alignment quality* 20; Mb: mega bases; pb: pares de bases.

As *reads* mapeadas cobriram toda a referência (aproximadamente 15,8 Kb de sequência) com uma alta cobertura – para fins de comparação, projetos de sequenciamento de genoma completo atingem em média 30X de profundidade de cobertura – e com boa densidade (90% da sequência referência apresentou cobertura superior a 100X) (TABELAS 5 e 6). A queda do percentual de cobertura é decorrente da perda de *reads* nas extremidades das áreas alvo. Prevendo este cenário, os *primers* foram desenvolvidos de maneira que a perda de informação seja de regiões flangeadoras (introns e UTRs), preservando assim áreas com maiores chances de abrigar variantes com maior potencial deletério (regiões codificantes e sítios de *splicing*).

TABELA 5. Número de *reads* e bases mapeados e suas respectivas coberturas (extensão) na referência.

Resumo dos alinhamentos processados	Controles	Pacientes
Número de <i>reads</i> mapeadas	353.086	193.783
Mapeamento de <i>reads</i> na referência	100%	100%
Número de bases mapeadas	52.825.640	25.953.543
Mapeamento de bases na referência	100%	100%

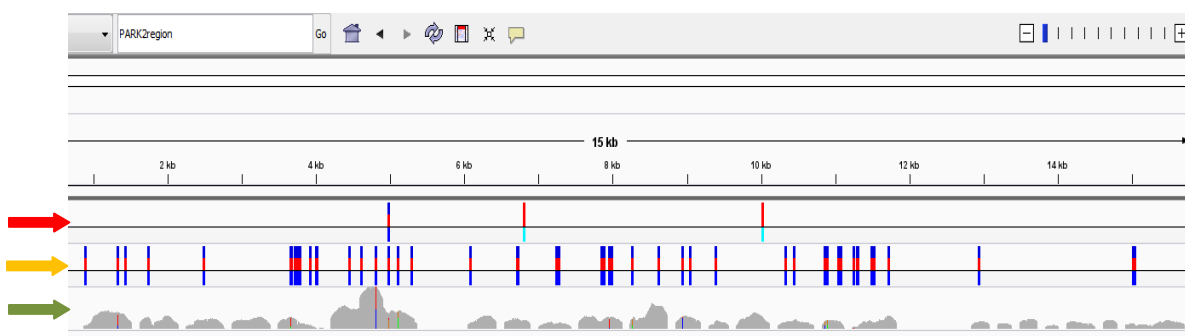
TABELA 6. Extensão da referência e dados de profundidade de cobertura.

Regiões alvos	Controles	Pacientes
Extensão da referência (pb)	15.805	15.805
Cobertura média	3.342,34	1.642,11
Uniformidade da cobertura	89,28%	89,44%
Cobertura a 1X	99,84%	98,56%
Cobertura a 20X	91,11%	90,66%
Cobertura a 100X	90,15%	90,04%

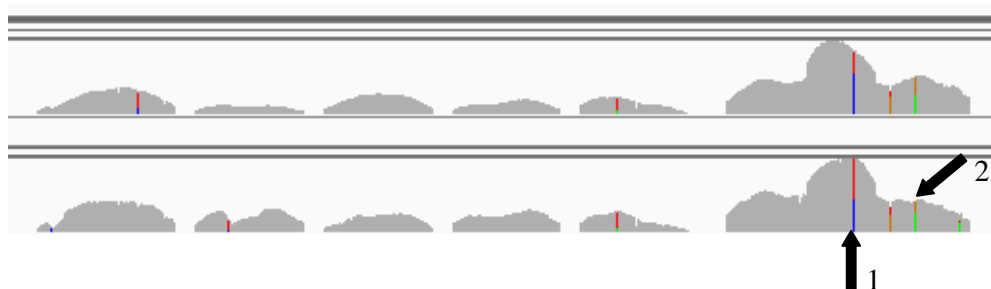
Pb: pares de bases.

Uma análise complementar ao TMAP foi realizada empregando o programa IGV que permite a visualização dos alinhamentos, conforme ilustrado na figura 10, que representa os 15,8 Kb de sequência referência e a cobertura de cada região alvo (FIGURA 10A). Observa-se que os genes *PACRG* (FIGURA 10B) e *PARK2* (FIGURA 10C) tiveram maior densidade de cobertura (até 5.000X em regiões do *PACRG* e até 4.000X no *PARK2*) quando comparados ao gene *QKI* (até 2.000X, FIGURA 10D). Esse fato deve-se ao tamanho menor dos amplicons do gene *QKI*, que podem ter sido perdidos em uma etapa de purificação realizada antes da concatenação. Nenhum dos genes estudados têm alinhamento nas regiões de exon 1, pois não foi possível padronizar a amplificação de um único amplicon destes exons nos três genes (o melhor experimento de PCR gerou 3 bandas de amplificação e o pior, 7).

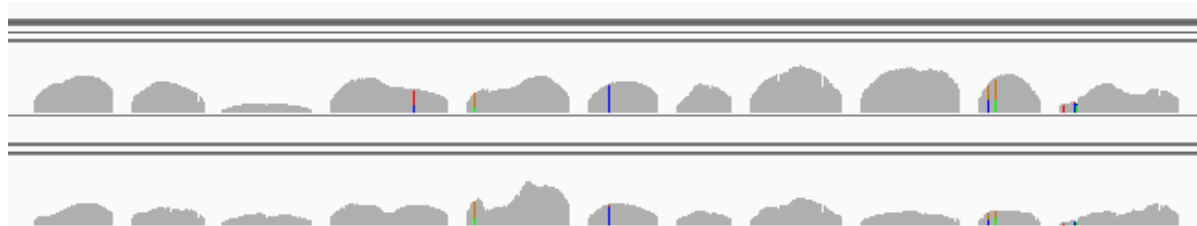
A



B



C



D

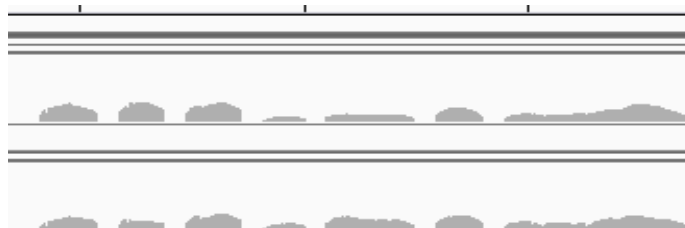


FIGURA 10. Visão esquemática do alinhamento e cobertura das regiões exônicas sequenciadas dos genes *PACRG*, *PARK2* e *QKI* segundo o algoritmo IGV. **A)** Visão geral do alinhamento e cobertura dos três genes em relação a uma sequência referência criada pela união das sequências referência das regiões exônicas. Seta vermelha: Campo sinalizador de indels. Seta amarela: Campo sinalizador de variações de ponto. Seta verde: Campo sinalizador de densidade de cobertura. **B)** Imagem ampliada do alinhamento de *PACRG* para o *pool* controles (em cima) e *pool* pacientes (em baixo). Setas pretas indicam a posição de variações de ponto, onde as linhas verticais coloridas indicam a posição de um SNV e a porcentagem de *reads* com determinada base. Código de cores: A, C, G, T. Seta 1: 50% C, 50% T. Seta 2: 70% A, 30% G. **C)** Imagem ampliada do alinhamento de *PARK2* para o *pool* controles (em cima) e *pool* pacientes (em baixo). **D)** Imagem ampliada do alinhamento de *QKI* para o *pool* controles (em cima) e *pool* pacientes (em baixo). Esta figura não inclui os exons 1 de cada gene.

A tabela 7 apresenta o número de variantes identificadas no sequenciamento e o número de variantes consideradas verdadeiras após o filtro de qualidade. O total de variantes identificadas foi de 82, sendo 45 variantes novas e 37 já descritas nas bases de dados *NCBI dbSNP*, *1000 Genomes* *HapMap*. No total, 38 variantes (46%) passaram pelo filtro de qualidade. Das variantes conhecidas, 84% foram aprovadas; o mesmo ocorreu para 16% das variantes novas. Este resultado é coerente, assumindo-se que: i) variantes já descritas e identificadas neste trabalho têm maior chance de serem reais e ii) variantes novas, principalmente de alto impacto, são menos prováveis de serem reais, dada a grande quantidade de trabalhos já publicados envolvendo a busca de variações causais de fenótipos da DP. Esses resultados argumentam a favor da eficiência dos filtros aplicados com o objetivo de redução de artefatos. O único gene que não teve nenhuma variante validada pelo filtro de qualidade foi o *QKI* (TABELA 7), resultado que indica um gene altamente conservado. O *PACRG* apresentou apenas 12 variantes validadas, três somente no *pool* controles e nove compartilhadas entre os *pools*. Já *PARK2* teve 26 variantes validadas, sendo sete presentes no *pool* pacientes, sete no *pool* controles, e 12 compartilhada entre os *pools*.

As tabelas 8, 9 e 10 apresentam as variantes conhecidas, conforme identificadas em ambos os *pools*, somente no *pool* controle ou somente no *pool* pacientes, respectivamente.

Das 23 variantes conhecidas detectadas em ambos os *pools*, 16 foram validadas porque atenderam todos os critérios de filtro de qualidade em ambos os *pools*, e três atingiram os critérios de validação apenas no *pool* controle (TABELA 8). Destas três variantes, duas foram excepcionalmente validadas também no *pool* pacientes apesar de não atingirem um dos critérios de qualidade. São elas, o SNP rs509495 que apresentou cobertura de 318X, frequência nas *reads* de 66,98%, razão Dir:Rev de 1,20, e Q_Var de 43, e o SNP rs6927767 que apresentou cobertura de 2027X, frequência nas *reads* de 4,49%, razão Dir:Rev de 2,11 e Q_Var de 48 (TABELA 8). Estas variantes não se encontravam em regiões de cobertura comprometida nem em *reads* com artefatos e homopolímeros. A variante rs199671842, apesar de atingir os critérios de qualidade no *pool* controles, não foi considerada verdadeira devido à presença de erros de sequenciamento nas *reads* em que a variante se encontrava.

TABELA 7. Número de variantes novas e já descritas encontradas nos *pool*/sem estudo para *PACRG*, *PARK2* e *QKI*.

Variantes	Filtro de qualidade	<i>PACRG</i>			<i>PARK2</i>			<i>QKI</i>			Total
		<i>Pool</i> Pacientes	<i>Pool</i> Controles	Compartilhadas	<i>Pool</i> Pacientes	<i>Pool</i> Controles	Compartilhadas	<i>Pool</i> Pacientes	<i>Pool</i> Controles	Compartilhadas	
Conhecidas	Antes	0	2	10	5	7	12	0	0	1	37
	Depois	0	2 (100%)	7 (70%)	4 (80%)	6 (86%)	12 (100%)	0	0	0 (0%)	31 (82%)
Novas	Antes	5	4	8	9	6	7	1	2	3	45
	Depois	0 (0%)	1 (25%)	1 (13%)	4 (44%)	1 (17%)	2 (29%)	0 (0%)	0 (0%)	0 (0%)	7 (16%)
Total	Antes	5	6	18	14	13	19	1	2	4	81
	Depois	0 (0%)	3 (50%)	8 (44%)	8 (57%)	7 (54%)	14 (74%)	0 (0%)	0 (0%)	0 (0%)	40 (49%)

TABELA 8. Variantes identificadas já descritas em bancos de dados que são compartilhadas entre os *pools* de pacientes e controles. Em negrito encontram-se aquelas que passaram pelo filtro de qualidade; em vermelho, variantes que foram excepcionalmente validadas, apesar de não atingir um dos critérios dos filtros.

Posição*	Alelos (Ref/Var)	Gene	Localização	dbSNP	PoolPacientes				PoolControles			
					Cobertura	Frequência da variante (%)	Razão Dir:Rev	Q_Var	Cobertura	Frequência da variante (%)	Razão Dir:Rev	Q_Var
6:161781225	G/A	PARK2	Exon 11	rs1801334	1948	1,90	1,65	47	6088	8,03	1,51	54
6:161970231	T/A	PARK2	Intron 8	rs509495	318	66,98	1,20	43	763	77,33	1,60	82
6:161990304	C/T	PARK2	Intron 8	rs75008325	1097	1,73	1,45	36	4808	3,45	1,01	53
6:161990339	G/A	PARK2	Intron 8	rs10945756	1433	16,82	1,15	46	6352	17,84	1,22	52
6:161990483	C/T	PARK2	Intron 7	rs3765474	1237	39,85	1,02	45	5753	31,90	1,07	100
6:161990516	G/C	PARK2	Intron 7	rs3765475	1571	61,68	1,05	46	6036	53,02	1,10	52
6:162206892	T/C	PARK2	Exon 7	rs9456711	1734	2,42	1,13	47	8882	1,08	1,29	100
6:162622197	C/T	PARK2	Exon 4	rs1801474	2025	5,28	1,50	48	5973	15,12	1,26	53
6:162622304	A/G	PARK2	Intron 3	rs4709583	2151	84,52	1,36	53	5274	96,97	1,08	100
6:162683876	C/T	PARK2	Intron 2	rs7744798	1316	48,18	3,21	53	2142	35,48	2,58	52
6:162864280	C/T	PARK2	Intron 2	rs113698817	1970	1,83	1,42	47	4222	7,79	1,25	55
6:162864317	T/C	PARK2	Intron 2	rs2075923	1639	21,48	1,98	51	4293	34,80	1,13	100
6:163149055	A/C	PACRG	Intron 1	rs9347683	650	13,38	1,05	56	982	15,48	1,57	100
6:163149497	T/C	PACRG	Intron 1	rs2276201	1960	16,48	1,14	51	3708	26,83	1,00	100
6:163733793	T/A	PACRG	Intron 5	rs56761274	1840	25,92	1,03	52	2720	33,16	1,05	100
6:163734134	A/G	PACRG	Intron 6	rs60293592	153	16,99	7,50	42	126	23,81	2,80	41
6:163734145	C/T	PACRG	Intron 6	rs199925383	108	23,15	6,00	42	94	15,96	2,75	44
6:163735853	C/A	PACRG	Intron 6	rs57286197	3555	1,21	2,83	46	5829	4,92	1,03	100
6:163736016	C/T	PACRG	Exon 7	rs60454188	6603	1,21	1,21	52	12421	4,85	1,31	52
6:163736204	T/C	PACRG	Exon 7	rs9306	7942	45,79	1,27	52	11880	65,33	1,05	52
6:163736502	A/G	PACRG	Exon 7	rs6903758	3176	37,37	1,08	52	6722	47,41	1,08	100
6:163736687	G/A	PACRG	Exon 7	rs6927767	2027	4,49	2,11	48	4595	20,76	1,57	53
6:163991700	G/T	QKI	Intron 7	rs199671842	368	4,35	2,18	41	839	4,53	1,92	45

dbSNP: *Single Nucleotide Polymorphism Database*; Dir: direta; Q_Var: Qualidade de identificação de Variante; Ref: referência; Rev: reversa; Var: variante. * *GRCh37.p10*, ago 2012.

Na tabela 9, observa-se que das nove variantes já descritas conhecidas detectadas no *pool* controles, sete foram validadas porque atenderam os filtros de qualidade, uma foi considerada artefato e uma foi excepcionalmente validada apesar de não atingir um dos critérios. Esta última é o SNP rs1801582 que foi considerado verdadeiro apesar de apresentarrarrazão Dir:Rev de2,17, porque apresentou cobertura de 5784X, frequência nas *reads* de 2,98%, Q_Var de 53 e ausência de artefatos e homopolímeros em suas *reads* de origem.

TABELA 9. Variantes identificadas já descritas em bancos de dados presentes somente em controles. Em negrito encontram-se os alelos que passaram pelo filtro de qualidade; em vermelho, a variante que foi excepcionalmente validada apesar de não atingir um dos critérios dos filtros.

Posição*	Alelos (Ref/Var)	Gene	Localização	dbSNP	Cobertura	Frequência da variante (%)	Razão Dir:Rev	Q_Var
6:161771037	A/G	<i>PARK2</i>	Exon 12	rs62637702	1765	1,05	1,0	46
6:161771115	G/A	<i>PARK2</i>	Exon 12	rs61730194	1558	1,15	1,1	46
6:161771219	C/T	<i>PARK2</i>	Exon 12	rs149953814	1537	1,47	1,35	41
6:161807855	C/G	<i>PARK2</i>	Exon 10	rs1801582	5784	2,98	2,17	53
6:161970171	A/C	<i>PARK2</i>	Intron 8	rs75976935	1014	7,40	4,2	43
6:161990344	A/G	<i>PARK2</i>	Intron 8	rs77112335	6473	1,17	1,03	52
6:161990401	C/T	<i>PARK2</i>	Exon 8	rs138920699	6599	1,36	1,45	53
6:163733891	C/A	<i>PACRG</i>	Exon 6	rs140329577	1987	1,64	1,69	36
6:163736266	T/C	<i>PACRG</i>	Exon 7	rs56157187	8856	1,05	1,13	100

dbSNP: *Single Nucleotide Polymorphism Database*; Dir: direta; Q_Var: Qualidade de identificação de Variante; Ref: referência; Rev: reversa; Var: variante. * *GRCh37.p10*, ago 2012.

Das cinco variantes já descritas que foram identificadas somente no *pool* pacientes, quatro foram consideradas verdadeiras porque atenderam os filtros de qualidade (TABELA 10).

TABELA 10. Variantes identificadas já descritas em bancos de dados presentes somente em pacientes. Em negrito encontram-se os alelos que passaram pelo filtro de qualidade.

Posição*	Alelos (Ref/Var)	Gene	Localização	dbSNP	Cobertura	Frequência da variante (%)	Razão Dir:Rev	Q_Var
6:161969761	A/G	<i>PARK2</i>	Intron 9	rs115564174	2369	2,74	1,78	48
6:162206998	C/T	<i>PARK2</i>	Intron 6	rs9458372	1362	2,13	1,14	45
6:162475167	A/C	<i>PARK2</i>	Exon 5	rs9456735	1480	1,82	1,36	45
6:162864377	G/A	<i>PARK2</i>	Exon 2	rs75860381	1668	1,92	1,11	46
6:162864412	A/G	<i>PARK2</i>	Exon 2	rs148851677	1218	1,31	6,60	30

dbSNP: *Single Nucleotide Polymorphism Database*; Dir: direta; Q_Var: Qualidade de identificação de Variante; Ref: referência; Rev: reversa; Var: variante. * *GRCh37.p10*, ago 2012.

A tabela 11 apresenta as variantes não descritas em bases de dados e que são compartilhadas em ambos os *pools*. No *pool* controles, o SNV 6:163736377 G/T foi considerado verdadeiro apesar de apresentar a razão Dir:Rev acima do valor de corte (2,62), dada sua alta frequência nas *reads* (19,06%), cobertura de 1731, Q_Var de 48 e ausência de artefatos e homopolímeros em suas *reads* de origem.

Em ambos os *pools*, as variantes 6:163733859 G/A e 6:163733908 T/C foram consideradas artefatos apesar de atingirem os critérios de qualidade, devido à presença de erros de sequenciamento nas *reads* em que estas variantes se encontravam. Situação que se repete no *pool* controle para as variantes 6:163733919 A/G, 6:163734058 T/G, 6:163991686 C/T e 6:163991710 G/T (TABELA 11).

TABELA 11. Variantes novas que são compartilhadas entre pacientes e controles. Em negrito encontram-se os alelos que passaram pelo filtro de qualidade; em vermelho, a variante que foi excepcionalmente validada apesar de não atingir um dos critérios dos filtros.

Posição*	Alelos (Ref/Var)	Gene	Localização	Pool/Pacientes				Pool/Controles			
				Cobertura	Frequência da variante (%)	Razão Dir:Rev	Q_Var	Cobertura	Frequência da variante (%)	Razão Dir:Rev	Q_Var
6:161771035	G/A	PARK2	Exon 12	720	3,33	1,82	43	1612	1,80	1,22	40
6:161771045	G/A	PARK2	Exon 12	830	1,57	1,50	29	1690	1,54	2,10	33
6:161970173	A/C	PARK2	Intron 8	279	16,85	3,40	42	760	13,29	4,71	40
6:161970174	T/C	PARK2	Intron 8	323	9,91	2,10	41	875	8,11	3,44	39
6:161970174	T/A	PARK2	Intron 8	323	7,74	2,45	41	875	8,23	3,8	41
6:161970175	C/T	PARK2	Intron 8	230	19,13	2,77	37	555	20,90	5,56	31
6:161970187	T/C	PARK2	Intron 8	345	2,61	5,75	30	726	6,06	5,3	30
6:163483223	G/T	PACRG	Exon 4	1267	2,13	15,14	45	2791	1,04	13,67	34
6:163733859	G/A	PACRG	Exon 6	1482	1,62	1,00	30	2360	1,06	1,26	31
6:163733897	A/G	PACRG	Exon 6	866	1,39	1,24	25	1566	1,47	1,33	25
6:163733908	T/C	PACRG	Exon 6	870	1,84	1,04	34	1377	2,03	1,24	34
6:163733919	A/G	PACRG	Intron 6	553	3,07	1,33	33	1214	1,89	1,33	31
6:163734058	T/G	PACRG	Intron 6	540	2,22	1,35	32	937	2,88	1,35	35
6:163734147	T/C	PACRG	Intron 6	102	6,86	2,33	33	82	6,10	2,0	33
6:163736377	G/T	PACRG	Exon 7	1056	23,58	1,89	51	1731	19,06	2,62	48
6:163991686	C/T	QKI	Intron 7	435	4,14	2,13	43	980	4,08	1,03	46
6:163991695	G/T	QKI	Intron 7	432	3,94	2,08	44	963	3,74	3,50	40
6:163991710	G/T	QKI	Intron 7	326	12,27	2,41	43	862	3,48	1,21	44

Dir: direta; Q_Var: Qualidade de identificação de Variante; Ref: referência; Rev: reversa; Var: variante. * GRCh37.p10, ago 2012.

A tabela 12 mostravariantes novas únicas de controles. Neste *pool*, os critérios de inspeção permitiram a seleção de apenas dois SNVs como bons candidatos. A variante 6:163734037 A/C foi consierada artefato por estar presente em *reads* com erros de sequenciamento. Em relação às variantes 6:162207037 T/A e 6:162207036 A/G, apesar de atingirem bons critérios de qualidade, ambas foram excluídas porque tinham origem nas mesmas *reads*, com muitos erros de sequenciamento e área de homopolímero.

TABELA 12. Variantes novas presentes somente em controles. Em negrito encontram-se os alelos que passaram pelo fitro de qualidade.

Posição*	Alelos (Ref/Var)	Gene	Localização	Cobertura	Frequência da variante (%)	Razão Dir:Rev	Q_Var
6:161970168	C/T	<i>PARK2</i>	Intron 8	1628	3,87	24,90	46
6:161970169	A/T	<i>PARK2</i>	Intron 8	1389	1,51	33,0	36
6:161970169	A/C	<i>PARK2</i>	Intron 8	1389	1,22	2,94	36
6:162207036	A/G	<i>PARK2</i>	Intron 6	7250	1,16	1,21	53
6:162207037	T/A	<i>PARK2</i>	Intron 6	7251	1,1	1,16	53
6:162207064	G/A	<i>PARK2</i>	Intron 6	6008	5,94	1,08	100
6:1631491980	A/T	<i>PACRG</i>	Intron 1	2456	1,38	16,40	44
6:163733925	T/C	<i>PACRG</i>	Exon 6	1695	1	1,28	25
6:163734037	A/C	<i>PACRG</i>	Intron 6	929	1,83	1,49	35
6:163736183	C/T	<i>PACRG</i>	Exon 7	11575	2,13	1,13	52
6:163876361	A/C	<i>QKI</i>	Exon 2	3669	1,44	15,00	53
6:163956151	A/T	<i>QKI</i>	Exon 4	3114	1,09	24,10	38

Dir: direta; Q_Var: Qualidade de identificação de Variante; Ref: referência; Rev: reversa; Var: variante. * GRCh37.p10, ago 2012.

A tabela 13 apresenta as variantes não descritas em bases de dados, presentes apenas no *pool* pacientes. De 15 variantes recém identificadas, somente três passaram pelo filtro de qualidade.

TABELA 13. Variantes novas presentes somente em pacientes. Em negrito encontram-se os alelos que passaram pelo fitro de qualidade.

Posição*	Alelos (Ref/Var)	Gene	Localização	Cobertura	Frequência da variante (%)	Razão Dir:Rev	Q_Var
6:161771052	C/T	<i>PARK2</i>	Exon 12	661	4,39	6,00	42
6:161781129 - 6:161781128	GA/G*	<i>PARK2</i>	Exon 11	986	18,15	1,52	100
6:161781207	C/T	<i>PARK2</i>	Exon 11	1875	3,15	1,89	48
6:161807734	A/G	<i>PARK2</i>	Intron 10	836	1,56	22,0	29
6:161969953	C/G	<i>PARK2</i>	Exon 9	969	3,2	37,67	44
6:161969964	T/G	<i>PARK2</i>	Exon 9	915	2,51	42,0	43
6:161969989	T/G	<i>PARK2</i>	Exon 9	1072	1,31	12,67	27
6:162394266 - 6:162394265	TC/T*	<i>PARK2</i>	Intron 6	986	20,99	1,14	100
6:162683517	G/C	<i>PARK2</i>	Intron 3	3847	2,16	2,81	55
6:163149592	A/G	<i>PACRG</i>	Intron 2	1875	1,01	11,8	27
6:163235134	A/T	<i>PACRG</i>	Intron 2	1668	4,5	18,43	46
6:163733804	G/T	<i>PACRG</i>	Intron 5	1723	1,51	28,50	39
6:163733852	A/T	<i>PACRG</i>	Exon 6	1642	1,1	1,0	28
6:163736374 - 6:163736377	CCT/C**	<i>PACRG</i>	Exon 7	1000	18,8	1,93	8
6:163876357	C/T	<i>QKI</i>	Exon 2	1381	1,67	2,67	39

Dir: direta; Q_Var: Qualidade de identificação de *Single Nucleotide Variant*; Ref: referência; Rev: reversa; Var: variante. * *GRCh37.p10*, ago 2012.

5.2.3 Análise terciária

5.2.3.1 Análise de impacto funcional

As variantes que passaram pelos critérios de inspeção foram analisadas quanto a sua presença em áreas de regulação de expressão (sítios de ligação de fatores transcricionais e áreas de regulação em UTRs), impacto em sítio de *splicing*, troca de aminoácido e, quando aplicável, análise de conservação de sequências protéicas e previsões de impacto funcional. As tabelas 14 e 15 apresentam os resultados dessas análises para as variantes já descritas em bases de dados e variantes inéditas, respectivamente.

Neste estudo foram detectadas seis variantes localizadas na região 3'UTR do gene *PACRG*, sendo quatro já descritas em bases de dados e duas novas, e três

variantes na região 3'UTR do *PARK2*, sendo duas já descritas e uma nova (TABELAS 14 e 15). Sabe-se que regiões 3' UTR são de grande importância para a estabilidade do RNAm; variantes nesta região podem diminuir ou abolir a função de um determinado RNAm (210). Por isso, estas variantes foram avaliadas quanto à possibilidade de modificação de sítios que alterem a dinâmica de transporte, tradução ou ligação de miRNA. Como resultado, não se encontrou evidência de envolvimento destas variantes com um comportamento aberrante dos RNAm em que estão presentes.

Na amostra estudada, foram detectadas 17 variantes localizadas em regiões intrônicas. Destas, seis são do gene *PACRG* (todas já descritas em bases de dados) e 11 do *PARK2* (nove já descritas e duas novas) (TABELAS 14 e 15). Estas variantes foram submetidas à análises *in silico* de previsão de alteração de sítios de *splicing* e sítios de ligação de miRNA, e nenhuma delas foi responsável por alterações nestes sítios. De acordo com a literatura e processo de anotação, o SNP rs9347683 é um polimorfismo de sítio de ligação de fator de transcrição.

Para o *PACRG* foram identificadas duas variantes conhecidas em regiões codificantes, uma delas é uma troca sinônima e a outra é uma variante missense. Já para o *PARK2* detectaram-se 10 variantes em regiões codificantes, oito delas já descritas e duas novas. Entre as variantes já descritas, duas são trocas sinônimas e seis são variantes missense (TABELA 14). E das duas variantes novas, uma é uma deleção que leva a uma mudança no quadro de leitura (*frameshift*) e a outra é um SNV que gera um códon de parada (*stop codon*) (TABELA 16). No item 5.4.3.3, as variantes de consequência funcional conhecida serão discutidas de acordo com suas previsões de impacto na proteína, sua localização na estrutura primária e presença em áreas conservadas do polipeptídeo.

TABELA 14. Variantes já descritas e filtradas: apresentação do tipo de variante, consequência funcional e previsão de impacto funcional.

Posição*	Gene	Localização	Tipo de variante	dbSNP	Base de dados**	Consequência funcional	Troca de aminoácido***	SIFT		PolyPhen-2	
								Score	Efeito	Score	Efeito
Variantes descritas compartilhadas pelos pools											
6:161781225	PARK2	Exon 11	SNV	rs1801334	1, 2, 3	Missense	394 D/N	0,023	Deletério	0,995	Provavelmente deletério
6:161970231	PARK2	Intron 8	SNV	rs509495	1, 2, 3	Desconhecida					
6:161990304	PARK2	Intron 8	SNV	rs75008325	1, 2	Desconhecida					
6:161990339	PARK2	Intron 8	SNV	rs10945756	1, 2, 3	Desconhecida					
6:161990483	PARK2	Intron 7	SNV	rs3765474	1, 2, 3	Desconhecida					
6:161990516	PARK2	Intron 7	SNV	rs3765475	1, 2, 3	Desconhecida					
6:162206892	PARK2	Exon 7	SNV	rs9456711	1, 2	Troca sinônima					
6:162622197	PARK2	Exon 4	SNV	rs1801474	1, 2, 3	Missense	167 S/N	0,155	Tolerável	0,012	Benigno
6:162622304	PARK2	Intron 3	SNV	rs4709583	1, 2	Desconhecida					
6:162864280	PARK2	Intron 2	SNV	rs113698817	1	Desconhecida					
6:162864317	PARK2	Intron 2	SNV	rs2075923	1, 2, 3	Desconhecida					
6:163149055	PACRG	Intron 1	SNV	rs9347683	1, 2, 3	Polimorfismo de sítio de ligação de fator de transcrição					
6:163149497	PACRG	Intron 1	SNV	rs2276201	1, 2, 3	Desconhecida					
6:163733793	PACRG	Intron 5	SNV	rs56761274	1, 2	Desconhecida					
6:163736016	PACRG	Exon 7	SNV	rs60454188	1, 2	Troca sinônima					
6:163736204	PACRG	3' UTR	SNV	rs9306	1, 2, 3	Desconhecida					
6:163736502	PACRG	3' UTR	SNV	rs6903758	1, 2, 3	Desconhecida					
6:163736687	PACRG	3' UTR	SNV	rs6927767	1, 2	Desconhecida					

(Continuação da tabela 14)

Variantes descritas únicas de controles											
6:161771037	PARK2	3' UTR	SNV	rs62637702	1,2	Desconhecida					
6:161771115	PARK2	3' UTR	SNV	rs61730194	1	Desconhecida					
6:161771219	PARK2	Exon 12	SNV	rs149953814	1	Missense	437 P/L	0,056	Tolerável	0,993	Provavelmente deletério
6:161807855	PARK2	Exon 10	SNV	rs1801582	1, 2, 3	Missense	380 V/L	0,377	Tolerável	0,001	Benigno
6:161990344	PARK2	Intron 8	SNV	rs77112335	1, 2	Desconhecida					
6:161990401	PARK2	Exon 8	SNV	rs138920699	1	Troca sinônima					
6:163733891	PACRG	Exon 6	SNV	rs140329577	1	Missense	232 P/T	0,701	Tolerável	0	Benigno
6:163735853	PACRG	Intron 6	SNV	rs57286197	1	Desconhecida					
6:163736266	PACRG	3' UTR	SNV	rs56157187	1, 2	Desconhecida					
Variantes descritas únicas de pacientes											
6:161969761	PARK2	Intron 9	SNV	rs115564174	1, 2	Desconhecida					
6:162206998	PARK2	Intron 6	SNV	rs9458372	1, 2	Desconhecida					
6:162475167	PARK2	Exon 5	SNV	rs9456735	1, 2, 3	Missense	192 M/L	0,055	Tolerável	0,001	Benigno
6:162864377	PARK2	Exon 2	SNV	rs75860381	1, 2	Missense	46 A/T	0	Deletério	1	Provavelmente deletério

* GRCh37.p10, ago 2012. ** 1: *dbSNP*; 2: *1000 genomes*; 3: *HapMap*. *** A: alanina; D: ácido aspártico; L: leucina; M: metionina; N: aspargina; P: prolina; PolyPhen-2: *Polymorphism phenotyping 2*; S: serina; SIFT: *Sorting intolerant from tolerant*; SNV: *Single Nucleotide Variant*; T: treonina; V: valina.

TABELA 15. Variantes novas filtradas e predição de impacto funcional.

Posição*	Gene	Localização	Tipo de variante	Consequência funcional
Variantes novas compartilhadas pelos pools				
6:161771035	<i>PARK2</i>	3' UTR	SNV	Desconhecida
6:163736377	<i>PACRG</i>	3' UTR	SNV	Desconhecida
Variantes novas únicas de controles				
6:162207064	<i>PARK2</i>	Intron 6	SNV	Desconhecida
6:163736183	<i>PACRG</i>	3' UTR	SNV	Desconhecida
Variantes novas únicas de pacientes				
6:161781129 - 6:161781128	<i>PARK2</i>	Exon 11	Deleção	<i>Frameshit</i>
6:161781207	<i>PARK2</i>	Exon 11	SNV	<i>Stop codon</i>
6:162394266 - 6:162394265	<i>PARK2</i>	Intron 6	Deleção	Desconhecida

SNV: *Single Nucleotide Variant*. * *GRCh37.p10*, Ago 2012

5.2.3.2 Análise de conservação

Das análises de conservação de sequências polipeptídicas, observa-se que a estrutura primária da parquina é bastante conservada entre as espécies analisadas, principalmente em áreas de domínio funcional, e com poucas áreas de *gaps* (falhas de alinhamento devido a diferenças no número de aminoácidos). Este resultado aponta que a extensão do polipeptídeo é importante para um dobramento correto, possivelmente com impacto na estrutura tridimensional necessária para sua função bioquímica (FIGURA 11). O mesmo não é observado para a proteína PACRG, uma vez que a extremidade inicial tem diversos *gaps* e uma importante diferença na região entre os aminoácidos 204 a 236, onde apenas as espécies mais próximas (humano e chimpanzé) apresentam alinhamento (FIGURA 12). De maneira geral, o PACRG tem muitas áreas conservadas e que são consistentes com a posição de domínios funcionais previstos para esta proteína. Este domínios não foram incluídos na figura 12 por ainda serem previsões, inferidas por análise de homologias de sequências, que necessitam de validação por estudos de cristalografia.

<i>Homo</i>	-----MIVFVRFNSSHGFPVEVSDSTSIFQLKEVVA	31
<i>Pan</i>	-----MIVFVRFNSSHGFPVEVSDSTSIFQLKEVVA	31
<i>Mus</i>	-----MIVFVRFNSSYGFPVEVSDSTSILQLKEVVA	31
<i>Rattus</i>	-----MIVFVRFNSSYGFPVEVSDSTSIFQLKEVVA	31
<i>Gallus</i>	-----MIVFVRFNSSHGFPVELGLDASILQLKEAVA	31
<i>Danio</i>	-----MIVFVRFNSSHGFPVELEQGASVSELKEAVG	31
<i>Drosophila</i>	MSFIFKFIATFVRKMLELLQFGGKTLTHTLSIYVKTNTGKTLTVNLEPQWDIKNVKELVA	60

rs75860381
46 A/T

UBL

<i>Homo</i>	KRQGV PADQLRVIFAGKELRNDWTVQNC DL DQQSIVHIVQRP-WRKGQEM--NATGGDDP	88
<i>Pan</i>	KRQGV PADQLRVIFAGKELRNDWTVQNC DL DQQSIVHIVQRP-WRKGQEM--NATGGDDP	88
<i>Mus</i>	KRQGV PADQLRVIFAGKELPNHLTVQNC DL EQQSIVHIVQRP-RRRSKET--NASGGDEP	88
<i>Rattus</i>	KRQGV PADQLRVIFAGKELQNHLLTVQNC DL EQQSIVHIVQRP-QRKSHE--NASGGDKP	88
<i>Gallus</i>	QRQGV PADQLRVIFAGRELNDLTLQNC DL VQQSIVHIVQNL-QKNSDKD--ET-EDNHA	87
<i>Danio</i>	RLQGV QSDQLRVIFAGREL CNESTLQGC DL PEQSTVHIVLPP-STSAHRS--ELIQRR-	87
<i>Drosophila</i>	PQLGLQDDDKIIFAGKELSDATTIEQCDLGQQSVLHAIRLRPPVQRQKIQSATLEEEEP	120

<i>Homo</i>	RNAAGGCEREPQSLTRVDLSSSVLPGDSVGLAVILHTDSRKDSPPAGSPAGRSIYNSFFV	148
<i>Pan</i>	RNAAGGCEREPQSLTRVDLSSSVLPGDSVGLAVILHTDSRKDSPPAGSPAGRSIYNSFFV	148
<i>Mus</i>	QSTSEGSIWESRSLTRVDLSSHTLPVDSVGLAVILDTDSKRDSEAARGPV-KPTYNSFFI	147
<i>Rattus</i>	QSTPEGSIWEPRLTRVDLSSSHILPADSVGLAVILDTDSKSDSEAARGPAAKPTYHSFFV	148
<i>Gallus</i>	GGILKTLERVPESLTRIDLSSSILPSLSAGLAVILDTKEPNISPPS-EKSGAASYNFFV	146
<i>Danio</i>	-----LGSGMESLTRLDLSSSRQTASEGLAVILETEASRREDA-GHTGAKAHSSFFV	140
<i>Drosophila</i>	SLSDEA--SKPLNETL-----LDLQLESEERLNIT--DEERVRAKAHFFV	161

rs1801474
167 S/N

rs9456735
192 M/L

<i>Homo</i>	YCKGPCQ RVQPGKLRVQSGTCRQATLTLTQGPS CWDDVLI PNFM SGECQSPHC-----	201
<i>Pan</i>	YCKGPCQ RVQPGKLRVQSGTCGQATLTLTQGPS CWDDVLI PNFM SGECQSPHC-----	201
<i>Mus</i>	YCKGPCCHKVQPGKLRVQSGTCRQATLTLAQGPS CWDDVLI PNFM SGECQSPDC-----	200
<i>Rattus</i>	YCKGPCCHKVQPGKLRVQSGTCRQATLTLAQGPS CWDDVLI PNFM SGECQSPDC-----	201
<i>Gallus</i>	FCKNFCQAVKPGKLRVRCNECKQGTTLTARGPS CWDDVLI PNFIITGVCSQSPDC-----	199
<i>Danio</i>	FCKTVCKAIQPGKLRVRCCKCKQGTTLTLRGPSCWDDVLLPNFIHGVCSQGC-----	193
<i>Drosophila</i>	HC-SQCDKLCNGKLRVRCALCKGGAFTVHRDPECWDDVLKSRFI PGHCSELEVACVDNAA	220

RING0

<i>Homo</i>	-PGTSAEFFFFKCGAHPTS-DKETSVALHLIATNSRNITCITCTDVRSPLVLFQCNSRHVI	259
<i>Pan</i>	-PGTSAEFFFFKCGAHPTS-DKETSVALHLIATNSRNITCITCTDVRSPLVLFQCNSRHVI	259
<i>Mus</i>	-PGTRAEEFFKCGAHPTS-DKDT SVALNLIITSNRRSIPCIAC T DVRSPLVLFQCNSRHVI	258
<i>Rattus</i>	-PGTRAEEFFKCGAHPTS-DKDT SVALNLIITNSRSIPCIAC T DVRNPLVLFQCNSRHVI	259
<i>Gallus</i>	-SGNVAEFYFKCGAHPTT-DSET SVALNLVTNTRSGITCITCTDIRSPLVLFQCMHRHVI	257
<i>Danio</i>	-NGRLAEFYLCASHPTC-DNDT SVALDIMPNTRRVPCIACTDIMTPVLVFCQAERHVI	251
<i>Drosophila</i>	GDPPFAEFFFFKCAEHVSGGEKDFAAPLNLIKNNIKNVPLACTDVSDTVLVEPCASQHV	280

RING1

<i>Homo</i>	CLDCFHLYCVTRLNDRQFVHDPQLGYS L P C V A G C P N S L I K E L H H F R I L G E E Q Y N R Y Q Q Y G	319
<i>Pan</i>	CLDCFHLYCVTRLNDRQFVHDPQLGYS L P C V A G C P N S L I K E L H H F R I L G E E Q Y N R Y Q Q Y G	319
<i>Mus</i>	CLDCFHLYCVTRLNDRQFVHDAQLGYS L P C V A G C P N S L I K E L H H F R I L G E E Q Y T R Y Q Q Y G	318
<i>Rattus</i>	CLDCFHLYCVTRLNDRQFVHDAQLGYS L P C V A G C P N S L I K E L H H F R I L G E E Q Y N R Y Q Q Y G	319
<i>Gallus</i>	CLDCFHLYCVTMLNDRQFIYDPDLGYS L P C V A G C P D S L I K E V H H F R I L G E E Q Y N R Y Q R Y G	317
<i>Danio</i>	CLECFHLYCVSRLNERQFTQETLLGYS L P C A A G C P D S L I K E V H H F R V L G E E Q Y E R Y Q R Y A	311
<i>Drosophila</i>	CIDCFRHYCRSRLGERQFMHPHDFGYTLPCPAGCEHSFIETIHHFKLLTREYDRYQFA	340

<i>Homo</i>	AEECVLQMGGVLCPRPGCAGLLPEPDQRKVTCEGG-NGLCGGFVFCRDCKEAYHEGEC	378
<i>Pan</i>	AEECVLQMGGVLCPRPGCAGLLPEPDQRKVTCEGG-NGLCGGFVFCRDCKEAYHEGEC	378
<i>Mus</i>	AEECVLQMGGVLCPRPGCAGLLPEQGGQRKVTCEGG-NGLCGGFVFCRDCKEAYHEGDCD	377
<i>Rattus</i>	AEECVLQMGGVLCPRPGCAGLLPEQGGQRKVTCEGG-NGLCGGFVFCRDCKEAYHEGEC	378
<i>Gallus</i>	AEECVLQMGGLLCPTPSCAGALLPEPEVRKIVCEPG-NGICGGFVFCRECKEYHEGEC	376
<i>Danio</i>	AEECVLQMGGVLCPPGCGAGLLPPDEERRVCCEPGNNGLCGGFVFCRDCKEEFHEGPCR	371
<i>Drosophila</i>	TEEYVLQAGGVLCPPGCGMGLLVEPDQRKVTQNG----CGYVFCRNCLQGYHIGEC	395

IBR

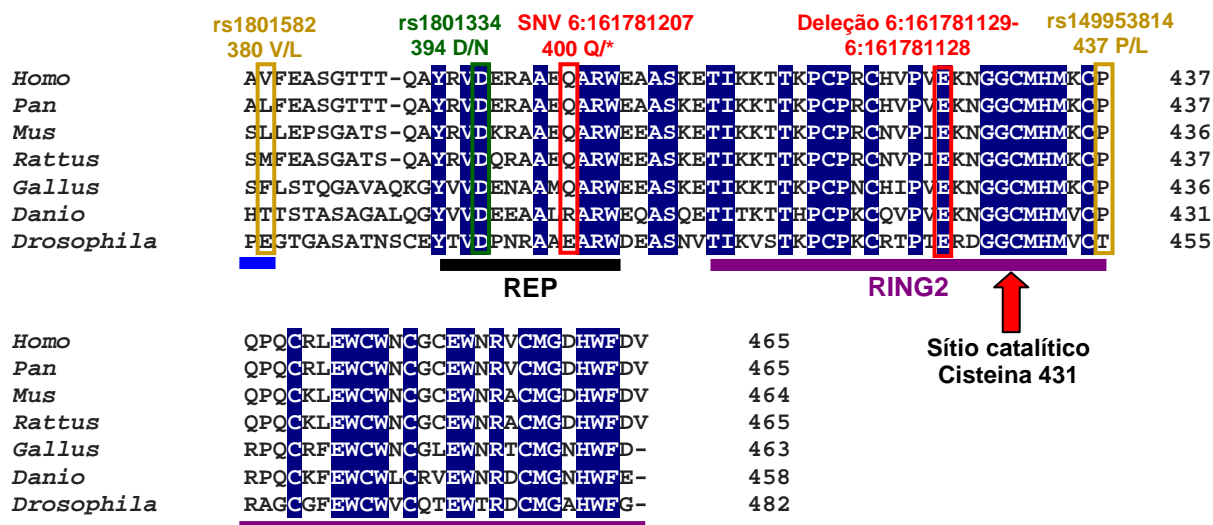


FIGURA 11. Análise de conservação das sequências de aminoácidos da proteína parkina entre as espécies *Homo sapiens*, *Pan troglodytes*, *Mus musculus*, *Rattus norvegicus*, *Gallus gallus*, *Danio rerio* e *Drosophila melanogaster*. Os aminoácidos conservados entre todas as espécies analisadas estão destacados em azul. As linhas horizontais representam os domínios da proteína. As variantes detectadas neste estudo estão apresentadas nos retângulos: em verde são as variantes detectadas em ambos os *pools*, em amarelo são as variantes detectadas somente no *pool* controles e em vermelho são as variantes detectadas somente no *pool* pacientes. A: alanina; C: cisteína; D: ácido aspártico; E: ácido glutâmico; F: fenilalanina; G: glicina; H: histidina; I: isoleucina; IBR: *In-Between-RING*; K: lisina; L: leucina; M: metionina; N: asparagina; P: prolina; Q: glutamina; R: arginina; REP: *Repressor Element of Parkin*; RING: *Really Interesting New Gene*; S: serina; T: treonina; UBL: *Ubiquitin-like*; V: valina; W: triptofano; Y: tirosina.

FONTE: Sequências de referência: NCBI (148-150); programa de alinhamento: Clustal Omega (206); domínios da proteína parkina: Trempe, 2013 (211).

<i>Homo</i>	--MVAEKETLSLNK--CPDKMPKRTK--L-LAQQP-LPVHQP----HSLVSEGETVKAMM	48
<i>Pan</i>	--MVAEKETLSLNK--CPDKMPKRTK--L-LAQQP-LPVHQP----HSLVSEGETVKAMM	48
<i>Mus</i>	-----MPKRTK--L-LPQQT-FQVHQP----RSLVSEGETVKAMM	32
<i>Rattus</i>	--MVAEKETLTLNK--CPDKMPKRTK--L-LPQQT-IQVHQP----HSLVSEGETVKAMM	48
<i>Gallus</i>	--MVVEE-----AG--CGPAKPAGNR--RPQKQEPLGHVKKT--KQQVSDGETVKAMM	45
<i>Danio</i>	-----MRT-FEPLAK-----GELKTQGETVMSTM	23
<i>Drosophila</i>	MAMAQTARTATARRPTHYHRPTRSKSANPAQLRPLSGIHGAHVSSRPYVPPESIQSQQ	60
<i>Homo</i>	KNSVVRGPPAA-----GAFKERPTKPTAFRKF-----	75
<i>Pan</i>	KNSVVRGPPAA-----GAFKERPTKPTAFRKF-----	75
<i>Mus</i>	KNSVVRGPPVA-----GAFKERPAKPTTFRK-----	58
<i>Rattus</i>	KNSVVRGPPAA-----GAFKERPTKPTAFRKFY-ERGDFFIALEHDSKGNKIAWKVRGP	101
<i>Gallus</i>	KNTVVRGPPLA-----GAFKERPTKP-----	66
<i>Danio</i>	KNSVVVGPPAA-----GAFREERPAK-----	43
<i>Drosophila</i>	KNTVVIDGPIHETAPKTASARSVPNPKILRRQKSMSTFN-----LGMGLNGC	109
<i>Homo</i>	-----YERGDFFIALEHDSKGNKIAWKVEIEKLDYHHYLPPLFFDGL	116
<i>Pan</i>	-----YERGDFFIALEHDSKGNKIAWKVEIEKLDYHHYLPPLFFDGL	116
<i>Mus</i>	-----CYERGDFFIALEHDSKGNKIAWKVEIEKLDYHHYLPPLFFDGL	100
<i>Rattus</i>	PAAGAFKERPTKPTAFRKFYERGDFFIALEHDSKGNKIAWKVEIEKLDYHHYLPPLFFDGL	161
<i>Gallus</i>	-----TAFRKFYERGDFFIALEHDTKGNRIAWKVEIEKLDYHHYLPPLFFDGL	113
<i>Danio</i>	-----PTAFRKFYERGDFFIALEHDSKGNRIAWKVEIEKLDYHHYLPPLFFDGL	91
<i>Drosophila</i>	ST--GGANDPGRGTLFRMYFDRGDLPIKMEYLCGDKIGWTVDIEKLDYSLYLPPLFFDGL	167
<i>Homo</i>	CEMTFPYEFFARQGIHDMLEHGGNKILPVLQQLIPIKNALNLRNRQVICVTLKVLQHLV	176
<i>Pan</i>	CEMTFPYEFFARQGIHDMLEHGGNKILPVLQQLIPIKNALNLRNRQVICVTLKVLQHLV	176
<i>Mus</i>	SEMTFPYEFFARRGIHDMLEHGGNKILPVIPLIPIKNALNLRNRQIICVTLKVLQHLV	160
<i>Rattus</i>	CEMTFPYEFFARQGIHDMLEHGGNKILPVIPLIPIKNALNLRNRQVICVTLKVLQHLV	221
<i>Gallus</i>	CEMTFPYEFFARQGIHDMLEHGGNKILPVIPLIPIKNALSLNRNRQVICVTLKVLQHLV	173
<i>Danio</i>	CETVHPYEFFARQGIHDMLEHGGNKVLPVIPLIPIKNALNTRNRQVICVTLKVLQHLV	151
<i>Drosophila</i>	AETKHPYKTYARQGVTDLLLAGGEKIHPVIPLIPLKNALSTRNLEVMCTTLKIQQQLV	227
	rs140329577 232 P/T	
<i>Homo</i>	VSAEMVGKALVPYYRQILPVLNIFKNMNGSYSLPRLECSGAIMARCNDLHGLSSDPPTSA	236
<i>Pan</i>	VSAEMVGKALVPYYRQILPVLNIFKNMNGSYSLPRLECSGAIMARCNDLHGLSSDPPTSA	236
<i>Mus</i>	VSSEMVGKALVPYYRQILPILNIFKNMN-----	188
<i>Rattus</i>	VSAEMVGKALVPYYRQILPILNIFKNMN-----	249
<i>Gallus</i>	VSADMVGKALVPYYRQILPVLNIFKNMN-----	201
<i>Danio</i>	VSAEMVGKALVPYYRQILPILNIFKNMN-----	179
<i>Drosophila</i>	MSSDLVGKALVPFYRQLLPENAFKVKN-----	255
<i>Homo</i>	SQVAEIIVNSGDGIDYSQQKRENIGDLIQETLEAFERYGGEDAFINIKYVVPPTYESCLLN	296
<i>Pan</i>	SQVAEIIVNSGDGIDYSQQKRENIGDLIQETLEAFERYGGEDAFINIKYVVPPTYESCLLN	296
<i>Mus</i>	-----VNSGDGIDYSQQKRENIGDLIQETLEAFERYGGEDAFINIKYVVPPTYESCLLN	241
<i>Rattus</i>	-----VNSGDGIDYSQQKRENIGDLIQETLEAFERYGGEDAFINIKYVVPPTYESCLLN	302
<i>Gallus</i>	-----VNSGDGIDYSQQKRENIGDLIQETLEAFERYGGEDAFINIKYVVPPTYESCLLN	254
<i>Danio</i>	-----KNSGDGIDYSQQKRENIGDLIQETLEAFERYGGEDAFINIKYVVPPTYESCLLN	232
<i>Drosophila</i>	-----LNCGEIDYAKNNLNLGDLIDETLQVLELHGGEDAFINIKYVVPPTYESCLLN	308

FIGURA 12. Análise de conservação das sequências de aminoácidos da proteína PACRG entre as espécies *Homo sapiens*, *Pan troglodytes*, *Mus musculus*, *Rattus norvegicus*, *Gallus gallus*, *Danio rerio* e *Drosophila melanogaster*. Os aminoácidos conservados entre todas as espécies analisadas estão destacados em azul. A variante detectada no pool controles está apresentada no retângulo amarelo. A: alanina; C: cisteína; D: ácido aspártico; E: ácido glutâmico; F: fenilalanina; G: glicina; H: histidina; I: isoleucina; K: lisina; L: leucina; M: metionina; N: asparagina; P: prolina; Q: glutamina; R: arginina; S: serina; T: treonina; V: valina; W: triptofano; Y: tirosina.

FONTE: Sequências de referência: NCBI (148-150); programa de alinhamento: Clustal Omega (206).

Já a proteína QKI mostrou se extremamente conservada nas espécies estudadas, exceto para *Drosophila melanogaster*, que não apresenta sequência homologa ao QKI (FIGURA 13). Além do domínio de ligação a RNAm (domínio KH), sabe-se que esta proteína contém motivos ricos em prolina, grupos tirosinas e repetições de arginina-glicina; motivos que são tipicamente encontrados em proteínas envolvida em processos de transdução de sinal (144). Estes dados sugerem que QKI tem participação em eventos celulares importantes; porém, precisamente quais e como esta proteína participa de rotas bioquímicas ainda é desconhecido (144). O alto grau de conservação do QKI é compatível com nosso achado de ausência de variantes neste gene. É provável que mutações deletérias em QKI possam ser incompatíveis com a vida, ao passo que a redução de sua expressão é tolerável, mas desencadeia fenótipos como aqueles observados na síndrome da deleção do cromossomo 6q (153).

<i>Homo</i>	MVGEMETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDDEEISVRKDMYNDT	60
<i>Pan</i>	-----MYNDT	5
<i>Mus</i>	MVGEMETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDDEEISVRKDMYNDT	60
<i>Rattus</i>	MVGEMETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDDEEISVRKDMYNDT	60
<i>Gallus</i>	MVGEMEAKKPKPSPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDDEEISVRKDMYNDT	60
<i>Danio</i>	MVGEMETKEKPKPTPDYLMQLMNDKKLMSSLPNFCGIFNHLERLLDDEEISVRKDMYNDT	60
<i>Homo</i>	LNGSTEKRSSELPLDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCK	120
<i>Pan</i>	LNGSTEKRSSELPLDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCK	65
<i>Mus</i>	LNGSTEKRSSELPLDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCK	120
<i>Rattus</i>	LNGSTEKRSSELPLDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCK	120
<i>Gallus</i>	LNGSTEKRSSELPLDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCK	120
<i>Danio</i>	LNGSTEKRSSELPLDAVGPIVQLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCK	120
<i>Homo</i>	IMVRGKGSMDKKKEEQNRGKPNWEHLNEDLHVLI TVEDAQNRAEIKLKRAVEEVKKLLV	180
<i>Pan</i>	IMVRGKGSMDKKKEEQNRGKPNWEHLNEDLHVLI TVEDAQNRAEIKLKRAVEEVKKLLV	125
<i>Mus</i>	IMVRGKGSMDKKKEEQNRGKPNWEHLNEDLHVLI TVEDAQNRAEIKLKRAVEEVKKLLV	180
<i>Rattus</i>	IMVRGKGSMDKKKEEQNRGKPNWEHLNEDLHVLI TVEDAQNRAEIKLKRAVEEVKKLLV	180
<i>Gallus</i>	IMVRGKGSMDKKKEEQNRGKPNWEHLNEDLHVLI TVEDAQNRAEIKLKRAVEEVKKLLI	180
<i>Danio</i>	IMVRGKGSMDKKKEEQNRGKPNWEHLNEDLHVLI TVEDAQNRAEIKLKRAVEEVNKKLLV	180
<i>Homo</i>	PAAEGEDSLKKMQLMELAILNGTYRDANIKSPALAFSLAATAQAAPRIITGPAPVLPPEAA	240
<i>Pan</i>	PAAEGEDSLKKMQLMELAILNGTYRDANIKSPALAFSLAATAQAAPRIITGPAPVLPPEAA	185
<i>Mus</i>	PAAEGEDSLKKMQLMELAILNGTYRDANIKSPALAFSLAATAQAAPRIITGPAPVLPPEAA	240
<i>Rattus</i>	PAAEGEDSLKKMQLMELAILNGTYRDANIKSPALAFSLAATAQAAPRIITGPAPVLPPEAA	240
<i>Gallus</i>	PAAEGEDSLKKMQLMELAILNGTYRDANIKSPALAFSLAATAQAAPRIITGPAPVLPPEAA	239
<i>Danio</i>	PAAEGEDSLKKMQLMELAILNGTYRDANIKSPALAFSLAASQAAPRIITGPAPVLPPEAA	239
<i>Homo</i>	LRTPTPAGPTIMPLIRQIQTA VMPNGTPHPT-AAIVPPGPEAGLIY-TPYEYPYTLAPAT	298
<i>Pan</i>	LRTPTPAGPTIMPLIRQIQTA VMPNGTPHPT-AAIVPPGPEAGLIY-TPYEYPYTLAPAT	243
<i>Mus</i>	LRTPTPAGPTIMPLIRQIQTA VMPNGTPHPT-AAIVPPGPEAGLIY-TPYEYPYTLAPAT	298
<i>Rattus</i>	LRTPTPAGPTIMPLIRQIQTA VMPNGTPHPT-AAIVPPGPEAGLIY-TPYEYPYTLAPAT	298
<i>Gallus</i>	LRTPTPAGPTIMPLIRQIQTA VMPNGTPHPT-AAIVPPGPEAGLIY-TPYEYPYTLAPAT	297
<i>Danio</i>	LRTPTPAGPTIMPLIRQIQTA VMPNGTAHPATATVPPGPESGLIYA TPYEYPYTLAPAT	298
<i>Homo</i>	SILEYPIEPSGVL EWIEMPVMPD-ISAH-----	325
<i>Pan</i>	SILEYPIEPSGVL EWIEMPVMPD-ISAH-----	270
<i>Mus</i>	SILEYPIEPSGVL EWIEMPVMPD-ISAH-----	325
<i>Rattus</i>	SILEYPIEPSGVL EWIEMPVMPD-ISAH-----	325
<i>Gallus</i>	SILEYPIEPSGVL GAVATKVRRHDMRVHPYQRIVTADRAATGN	340
<i>Danio</i>	SILEYPIDSSGVL GVAFPPTKG-----	319

FIGURA 13. Análise de conservação das sequências de aminoácidos da proteína QKI entre as espécies *Homo sapiens*, *Pan troglodytes*, *Mus musculus*, *Rattus norvegicus*, *Gallus gallus*, *Danio rerio*. Os aminoácidos conservados entre todas as espécies analisadas estão destacados em azul. A: alanina; C: cisteína; D: ácido aspártico; E: ácido glutâmico; F: fenilalanina; G: glicina; H: histidina; I: isoleucina; K: lisina; L: leucina; M: metionina; N: asparagina; P: prolina; Q: glutamina; R: arginina; S: serina; T: treonina; V: valina; W: triptofano; Y: tirosina.

FONTE: Sequências de referência: NCBI (148-150); programa de alinhamento: Clustal Omega (206).

5.2.3.3 Interpretação dos resultados

5.2.3.3.1 *PACRG*

A variante rs9347683 A/C, compartilhada entre os *pools*, é um polimorfismo de sítio de ligação de fatores transcricionais localizado no intron 1 do gene *PACRG*, cuja posição é a -258 em relação ao gene *PARK2*. Esta variante encontra-se, portanto, na sequência promotora do *PARK2* e já foi descrita em associação com DP por dois trabalhos publicados em 2002. Aplicando a técnica de sequenciamento de Sanger, o primeiro grupo descreveu a variante em heterozigose como sendo responsável por DPIP em 61% de 18 indivíduos numa amostra populacional norte-europeia (212). Meses mais tarde West *et al.* avaliaram este polimorfismo em uma amostra de 319 pacientes de DP clássica e 196 controles. Análise *in silico* realizada por estes pesquisadores indicou que o alelo alternativo C reduz a afinidade do sítio de ligação a fatores transcricionais, resultado reforçado por ensaio de atividade enzimática da proteína parkina *in vivo*, no qual o alelo alternativo foi responsável por uma redução de 25% de atividade em relação ao alelo referência (134). Estes trabalhos concluem que o SNP rs9347683 pode ser responsável pela redução da expressão de *PARK2*, desta forma predispondo o indivíduo à DP. Curiosamente esta variante foi relacionada tanto a DPIP como a forma clássica de DP e em indivíduos homozigotos ela parece influenciar na idade de início dos sintomas levando à forma de início precoce (213). Em contraste com os trabalhos citados, Ross *et al.* (214) e Taylor *et al.* (215) não identificaram relação entre o polimorfismo e DP. Trabalho mais recente conduzido em uma população chinesa de 595 casos e 492 controles também não encontrou associação entre o polimorfismo e pacientes tanto da forma precoce como tardia da doença (216).

Na nossa amostra populacional, o polimorfismo rs9347683 apresenta uma pequena diferença de distribuição de frequências *reads* entre os *pools*, sendo mais frequente no *pool* controle (13,38% e 15,48% para os *pools* pacientes e controles respectivamente - Tabela 8). Assim, é pouco provável que esta variante tenha significado biológico para a etiologia de DPIP na amostra populacional estudada.

Observado apenas no *pool* controle, o SNP rs140329577 C/A (232 P/T)

resulta em uma troca de prolina, um aminoácido apolar, por treonina, um aminoácido polar não carregado, na proteína PACRG. Esta variante encontra-se em uma área não conservada do gene (FIGURA 12), levando o seu impacto, em nossa análise, a ser classificado como tolerável (TABELA 14).

5.2.3.3.2 *PARK2*

O SNP rs75860381 G/A identificado apenas no *pool* paciente, está localizado no exon 2 do *PARK2* e gera a troca de um aminoácido apolar por um polar neutro (alanina por treonina) na posição 46 da proteína, uma região bastante conservada do domínio UBL (FIGURA 11). A previsão de impacto aponta esta variante como sendo muito deletéria (0 para SIFT e 1 para PolyPhen-2 - TABELA 14), pois ela se localiza em motivo altamente conservado (sequência IFAG), provavelmente envolvido na interação da parquína com proteínas E2 (211). Desta forma, é razoável supor que esta mutação pode levar à perda ou redução da atividade da parquína por impossibilitar a transferência de moléculas de ubiquitina para processamento. Apesar de considerada bastante deletéria à estrutura nativa da parquína, curiosamente não havia sido reportada por estudos em doença de Parkinson. No entanto, dada a localização desta variante na proteína, sua previsão de impacto e a ocorrência somente no *pool* paciente com uma frequência de 1,92%; conclui-se que esta variante possa responder por pelo menos um caso heterozigoto para a variante na amostra estudada.

A variante rs1801474 C/T (167 S/N), localizada no exon 4 do *PARK2* (troca de uma serina por uma asparagina - ambos polares e neutros), apesar de apresentar frequências de 5,28% em pacientes e 15,12% em controles (TABELA 8) e estar ao lado de uma cisteína conservada (C166, domínio RING0 - FIGURA 11) que coordena ligação eletrostática com íon de zinco, não aparenta ter qualquer influência no fenótipo estudado, uma vez que teve previsão de efeito tolerável (TABELA 14) e não mostrou-se conservado entre as espécies analisadas (FIGURA 11). Aparentemente, as estruturas *Zinc Finger* são bastante resistentes quanto a variações em aminoácidos que não coordenam ligações com o zinco. Um exemplo interessante é fornecido por mutações no aminoácido 161K, que faz parte da alça de

coordenação do *Zinc Finger* no domínio RING0, que não levam a alterações de atividade enzimática (211).

O polimorfismo rs9456735 A/C(192 M/L) encontrado apenas no *pool* paciente corresponde a uma troca de metionina por leucina, aminoácidos apolares e neutros, em uma posição pouco conservada do domínio RING0. Esta variante também está localizada próximo de uma cisteína conservada (C196) que coordena ligação eletrostática com íon de zinco. De acordo com os algoritmos de previsão de impacto, esta variante é considerada tolerável. Esta variante já está descrita nas bases de dados, no entanto uma outra variante com o mesmo registro dbSNP mas que resulta na troca 192M/V é que foi identificada em pacientes portadores de DPIP e paciente de DP tardio familiar (217, 218). Para a variante 192M/V, existem estudos *in vitro* que indicam associação desta variação com hipoatividade da proteína parquína, sendo mais recente o trabalho que resolveu com maiores detalhes a estrutura desta proteína (211). Os polimorfismos 192M/V e 192M/L estão presentes em uma área do domínio RING0 responsável por interagir diretamente com o domínio RING2. Neste ponto, há um controle de atividade enzimática por oclusão do sítio catalítico criada pelos dois domínios (211). Aparentemente, a ocorrência de V192 deve promover mudanças topológicas e/ou eletrostáticas que são desfavoráveis a exposição do sítio catalítico, dada a grande diferença de estrutura entre os aminoácidos metionina e valina. Essa situação parece não ser tão drástica com a troca L192, uma vez que a leucina se assemelha em volume (extensão da molécula) à metionina, portanto, a previsão é de impacto tolerável.

Exclusiva das *reads* do *pool* controle em frequência de 2,98%, a variante rs1801582 C/G (380V/L) no exon 10 do *PARK2*, que consiste na troca de valina por leucina (ambos AA apolares), foi classificada como tolerável (TABELA 14), uma vez que a valina não é conservada e é o último AA do domínio IBR (FIGURA 11).

O SNP rs1801334 G/A (394 D/N), localizado no exon 11, corresponde a uma troca do AA aspartato (de característica polar e ácido) por asparagina (um AA polar neutro). Este polimorfismo está na região repressora de atividade da proteína parquína (FIGURA 11), uma estrutura em α -hélice responsável pelo controle de interação da parquína, via impedimento estérico, com o sítio de ligação às proteínas E2 (211). Análise de previsão de impacto indica que a variante é danosa à conformação nativa da proteína (TABELA 14); Compatível com esta previsão é a

observação de que o aminoácido original em questão é bastante conservado em diferentes espécies (FIGURA 11). Além disso, a troca de classes de aminoácido pode desencadear alteração de interações normais entre os AA originais e causar instabilidade na proteína. Estas informações permitem sugerir que o AA nativo exerce importante função de estabilização da região repressora por interações eletrostáticas com AA vizinhos. Caso a diferença de frequência observada nas *reads* entre os grupos de pacientes e controles (1,90% e 8,03%, respectivamente – tabela 8) reflita a frequência real deste SNP entre estes grupos, a variante pode estar relacionada a um efeito de proteção contra a ocorrência de DP, uma vez que (i) este SNP é quatro vezes mais frequente nas *reads* do *pool* controles nesta amostra populacional e (ii) pela demonstração de que variantes em áreas que desestabilizam o repressor induzem uma maior atividade da enzima (211, 219). Sabe-se que parkina tem função na homeostase celular tanto por marcar proteínas velhas ou disfuncionais para degradação quanto por promover autofagia de mitocôndrias danificadas (220). Desta forma, o SNP rs1801334 G/A pode estar causando um menor limiar de ativação de PARK2, tornando-a mais efetiva, por exemplo, em vias de promoção de mitofagia, recuperando neurônio sem estágios iniciais de apoptose.

Os SNPs rs1801474 (167S/N), rs1801582 (380V/L) e rs1801334 (394D/N) identificados neste trabalho já foram estudados em DP; no entanto, na literatura apontados resultados divergentes.

Lucking *et al.* (221) avaliaram o efeito destas variantes em uma amostra populacional europeia na DP familiar e na DP esporádica. Como resultado, não foi detectada associação da variante S167N com nenhum dos tipos de DP estudados. Por outro lado, o polimorfismo V380 em homozigose apresentou-se em associação com DP em um grupo de portadores de doença esporádica. No entanto, este efeito foi independente de idade de início da doença. Já o polimorfismo D394 em homozigose apresentou tendência de associação com DP familiar.

Em 2012, Gaweda-Walerych, *et al.* (222) genotiparamas variantes 167 S/N, 380 V/L e 394 D/N em uma amostra populacional polonesa de portadores de DPIP, porém não encontraram relação causal entre as três variantes e manifestação de DPIP. No entanto, ao fazer uma análise refinada da co-ocorrências destas variantes com haplogrupos específicos de mitocôndrias e genótipos do gene *TFAM*, eles identificaram um resultado interessante: a co-ocorrência do genótipo rs2306604 G/G

do *TFAM*, com a variante 380V/L em indivíduos portadores do haplogrupo mitocondrial HV foi associada a um risco aumentado de desenvolvimento de DPIIP (OR: 3,94, IC 95%: 1,82–8,53, $P=0,0003$).

Em uma amostra norte-americana de 101 portadores de DP e 105 controles, Clark *et al.* sequenciaram todas as regiões exônicas de *PARK2* e identificaram as mesma três variantes, porém, não houve diferença estatisticamente significativa na frequência destas variantes entre os grupos (223).

Por último, o SNP rs149953814 C/T (437 P/L), localizado no exon 12 do *PARK2*, foi detectado somente nas *reads* do *pool* controle com frequência de 1,47%. Este polimorfismo gera uma substituição de uma prolina por leucina ao lado de C436 que participa de uma ligação de coordenação zinco. Apesar desta variante trocar aminoácidos de mesma classe (ambos apolares), a posição é bem conservada em 6 das 7 espécies analisadas e está em proximidade a C431, sítio catalítico de *PARK2* (211) (FIGURA 11). O programa SIFT aponta uma tolerabilidade limítrofe (no caso um score de 0,056, sendo que abaixo de 0,05 a previsão passa a ser deletéria), e o Polyphen-2 aponta uma característica provavelmente danosa (0,993) (TABELA 14). Esta variante aparenta ser deletéria para a estrutura nativa da parquína e, a exemplo da variante 192M/V, a substituição pode levar a uma hipofunção da enzima uma vez que a posição 437 da proteína é pouco variável entre as espécies, além de estar localizada em estreita proximidade ao sítio catalítico (FIGURA 11). No entanto a frequência na qual a variação 437 P/L aparece nas *reads* no *pool* controle permite inferir que ela esteja presente em apenas 1 cromossomo. Sendo uma variante do *pool* controle, é possível especular que um indivíduo heterozigoto para esta variante não está sob risco dedesenvolver DPIIP; no entanto, este indivíduo poderia ser acompanhado para avaliar-se o papel desta variante no desenvolvimento da DP clássica.

Embora promissores, o efeito das variantes discutidas acima precisam ser investigado em estudos adicionais, como por exemplo, ensaios celulares, para se confirmar um efeito biológico.

Das variantes novas identificadas, apenas duas, presentes somente nas *reads* do *pool* pacientes, são de potencial impacto funcional, ambas no gene *PARK2*. Se confirmadas, estas variantes podem responder por 21% dos casos de DPIIP na amostra pesquisada. O SNV 6:161781207 C/T, localizado na posição 31 do

exon 11 (códon que codifica 400Q na proteína parkina), induz uma mutação *nonsense* (400 Q/*) na região REP da proteína (FIGURAs 11 e 14C). Esta variante produz uma proteína truncada, que perde completamente o domínio RING2 que contém o sítio catalítico de atividade ubiquitina-ligase (211). Sabe-se que mutações *nonsense* são responsáveis por desencadear o mecanismo *nonsense mediated decay*, que faz com que RNA nascentes sejam degradados antes que possam ser traduzidos, ou mesmo distribuídos intracelularmente (224). Desta forma, esta variante é considerada de potencial grande impacto em dois cenários: em estado heterozigoto, pode levar a uma importante redução da expressão de *PARK2*, gerando um quadro de haploinsuficiência; em homozigose, pode levar à abolição de expressão e/ou atividade de *PARK2*, comprometendo as funções da parkina na vigília da homeostase celular, na reciclagem de proteínas danificadas e/ou na atividade promotora de autofagia de mitocôndrias danificadas. Esta variante pode ser responsável por 3% dos casos de DPIP presentes neste trabalho.

De consequência funcional potencialmente mais deletéria, a deleção 6:161781129 - 6:161781128 GA/G pode ser responsável por 18% dos casos de DPIP na amostra estudada. Esta mutação leva a uma mudança do quadro de leitura a partir da posição 110 (ou códon GAA) do exon 11 do gene *PARK2*. Esta deleção leva à modificação do códon 426, que codifica um resíduo de glutamato (aminoácido conservado e a 5 mer de distância do sítio catalítico C431 - FIGURA 11). Esta deleção desencadeia a alteração de toda a sequência de leitura adiante e o resultado é a tradução de uma proteína com 8 aminoácidos diferentes da original a partir do códon 426, sem o sítio catalítico C431 e truncada, uma vez que a mudança do quadro de leitura introduz um códon de parada precoce TGA (FIGURA 14B). Por induzir a perda completa do domínio RING2, esta mutação resulta na abolição da função de ubiquitina-ligase da parkina.

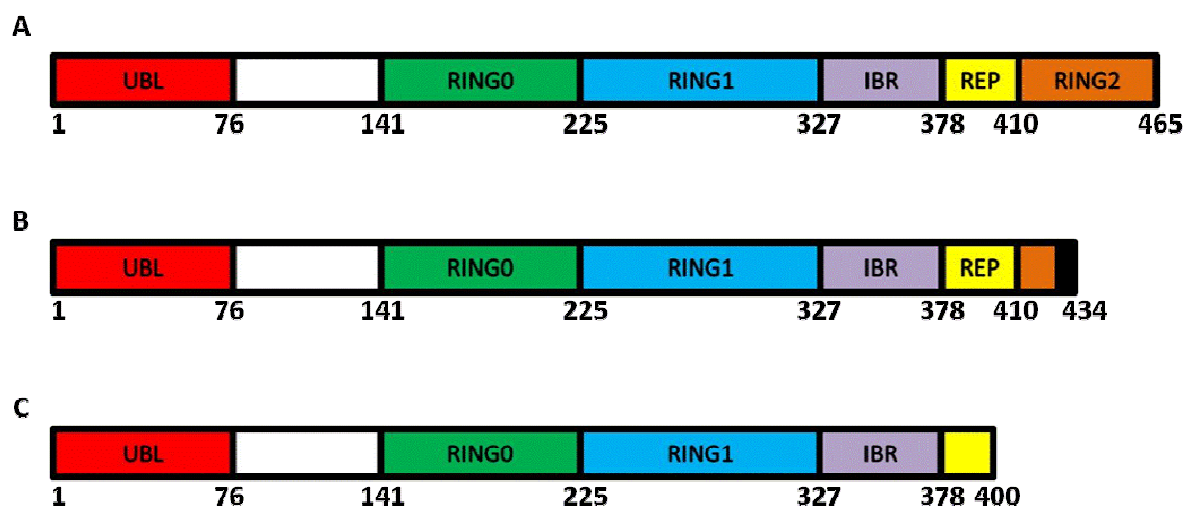


FIGURA 14. Apresentação da estrutura primária da parquína na sua forma nativa e alterada por duas mutações. **A)** Estrutura primária da parquína com seus domínios e extensão total; **B)** Consequência da deleção 6:161781129 - 6:161781128 GA/G, que induz um *frameshift*; **C)** Consequência da mutação nonsense SNV 6:161781207 C/T. IBR: *In-Between-RING*; REP: *Repressor Element of Parkin*; RING: *Really Interesting New Gene*; UBL: *Ubiquitin-like*
 FONTE: Adaptado de Treme, 2013(211)

5.3 LIMITAÇÕES

5.3.1 Limitações do estudo

Este trabalho tem duas limitações importantes, ambas superáveis através da mesma estratégia. Primeiro, o estudo sofre da ausência de experimento de validação das variantes identificadas, inéditas ou conhecidas, através de técnica de sequenciamento consagrada, como por exemplo, o método de Sanger baseado em fluorescência e eletroforese capilar. Uma segunda limitação está na impossibilidade de se definir quais indivíduos são os portadores das variantes de interesse; como consequência, a determinação exata das frequências alélicas e genotípicas fica comprometida.

Ambas as limitações são inerentes ao uso de uma técnica recente, ainda pouco explorada e em contínuo desenvolvimento, como a de sequenciamento de próxima geração implementada na plataforma *Ion PGM*. Para superá-las, estão em andamento no nosso laboratório duas estratégias complementares de análise: (i) o sequenciamento das amostras individuais para os *loci* contendo as variantes de

maior interesse, através de método de Sanger, e (ii) a geração das mesmas sequências aqui descritas, também por sequenciamento de próxima geração, agora produzidas em plataforma SOLiD, de natureza distinta à disponível no *Ion Torrent*. Esta validação é mandatória para o seguimento da investigação, e será realizada como condição para o preparo do manuscrito e submissão para publicação.

Uma terceira limitação é relativa ao pequeno número de indivíduos controle, a fim de se alcançar um melhor pareamento com o grupo de casos. Neste sentido, uma amostra expandida deste grupo controle já está recrutada e disponível para novos experimentos de sequenciamento.

6 CONCLUSÕES

- A aplicação de plataforma de sequenciamento profundo levou à identificação de 81 variantes candidatas a validação, distribuídas nas sequências codificadoras e flangeadoras imediatas dos genes *QKI*, *PACRG* e *PARK2*, em dois *pools* formados por DNA de pacientes portadores de DPIP e indivíduos livres de DPIP ou qualquer outra doença neurodegenerativa;
- Os filtros de qualidade desenvolvidos neste estudo levaram à sugestão de confirmação de 38 variantes; destas, seis apresentaram potencial impacto funcional;
- Foram identificadas 7 variantes candidatas em sequências do gene *QKI*, mas estas variantes não foram validadas pelos filtros de qualidade.
- Foi encontrada evidência de duas variantes funcionais no gene *PACRG*; uma (rs140329577 - p.P232T), presente apenas no pool controle, que tem previsão de impacto tolerável; a outra, rs9347683, é um polimorfismo de região promotora de *PARK2* que, apesar de ter sido demonstrado que sua ocorrência reduz a expressão de *PARK2*, sua associação com DP ou DPIP não foi consistentemente replicada;
- Foi encontrada evidência de quatro variantes causadoras de impacto funcional no gene *PARK2*, duas novas e duas já conhecidas. O rs75860381 - p.A46T, presente no *pool* paciente, tem previsão de impacto deletério e esta presente no domínio UBL da parquina, domínio de importância para interação com proteínas E2, portanto sendo uma variante de susceptibilidade. O polimorfismo rs1801334 - p.D394N é compartilhado pelos *pools* – porém apresenta frequência quatro vezes maior no pool paciente, encontra-se na região repressora de atividade da parquina e tem predição de impacto deletéria a estrutura nativa da proteína; esta substituição pode resultar em uma maior atividade da enzima resultando em efeito de proteção. As duas variantes novas identificadas neste estudo e presente apenas no *pool* paciente são a troca C/T na posição 6:161781207 que resulta em mutação *nonsense* que leva à perda do sítio catalítico da parquina; e a deleção GA/G na posição 6:161781129 - 6:161781128 que desencadeia um *frameshift* na área que codifica o sítio catalítico da parquina. Ambas as variantes são de alto

impacto e levam a uma completa abolição da função da parquina.

- Em conjunto, os resultados aqui presentes sugerem que o *PARK2* é o único responsável pela ligação do locus 6q25-q27 à DP1P.

REFERÊNCIAS

1. Parkinson J. An essay on the shaking palsy. 1817. The Journal of neuropsychiatry and clinical neurosciences. 2002 Spring;14(2):223-36; discussion 2. PubMed PMID: 11983801.
2. Lees AJ, Hardy J, Revesz T. Parkinson's disease. Lancet. 2009 Jun 13;373(9680):2055-66. PubMed PMID: 19524782.
3. de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. Lancet neurology. 2006 Jun;5(6):525-35. PubMed PMID: 16713924.
4. Bekris LM, Mata IF, Zabetian CP. The genetics of Parkinson disease. Journal of geriatric psychiatry and neurology. 2010 Dec;23(4):228-42. PubMed PMID: 20938043. Pubmed Central PMCID: 3044594.
5. Obeso JA, Rodriguez-Oroz MC, Goetz CG, Marin C, Kordower JH, Rodriguez M, et al. Missing pieces in the Parkinson's disease puzzle. Nature medicine. 2010 Jun;16(6):653-61. PubMed PMID: 20495568.
6. Barbosa MT, Caramelli P, Maia DP, Cunningham MC, Guerra HL, Lima-Costa MF, et al. Parkinsonism and Parkinson's disease in the elderly: a community-based survey in Brazil (the Bambui study). Movement disorders : official journal of the Movement Disorder Society. 2006 Jun;21(6):800-8. PubMed PMID: 16482566.
7. Wooten GF, Currie LJ, Bovbjerg VE, Lee JK, Patrie J. Are men at greater risk for Parkinson's disease than women? Journal of neurology, neurosurgery, and psychiatry. 2004 Apr;75(4):637-9. PubMed PMID: 15026515. Pubmed Central PMCID: 1739032.
8. Twelves D, Perkins KS, Counsell C. Systematic review of incidence studies of Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society. 2003 Jan;18(1):19-31. PubMed PMID: 12518297.
9. Van Den Eeden SK, Tanner CM, Bernstein AL, Fross RD, Leimpeter A, Bloch DA, et al. Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. American journal of epidemiology. 2003 Jun 1;157(11):1015-22. PubMed PMID: 12777365.
10. Schrag A, Schott JM. Epidemiological, clinical, and genetic characteristics of

early-onset parkinsonism. *Lancet neurology*. 2006 Apr;5(4):355-63. PubMed PMID: 16545752.

11. Gershanik OS, Leist A. Juvenile onset Parkinson's disease. *Advances in neurology*. 1987;45:213-6. PubMed PMID: 3825691.
12. Lucking CB, Durr A, Bonifati V, Vaughan J, De Michele G, Gasser T, et al. Association between early-onset Parkinson's disease and mutations in the parkin gene. *The New England journal of medicine*. 2000 May 25;342(21):1560-7. PubMed PMID: 10824074.
13. Periquet M, Latouche M, Lohmann E, Rawal N, De Michele G, Ricard S, et al. Parkin mutations are frequent in patients with isolated early-onset parkinsonism. *Brain : a journal of neurology*. 2003 Jun;126(Pt 6):1271-8. PubMed PMID: 12764050.
14. Jankovic J. Parkinson's disease: clinical features and diagnosis. *Journal of neurology, neurosurgery, and psychiatry*. 2008 Apr;79(4):368-76. PubMed PMID: 18344392.
15. Fearnley JM, Lees AJ. Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain : a journal of neurology*. 1991 Oct;114 (Pt 5):2283-301. PubMed PMID: 1933245.
16. Shulman JM, De Jager PL, Feany MB. Parkinson's disease: genetics and pathogenesis. *Annual review of pathology*. 2011;6:193-222. PubMed PMID: 21034221.
17. Meireles J, Massano J. Cognitive impairment and dementia in Parkinson's disease: clinical features, diagnosis, and management. *Frontiers in neurology*. 2012;3:88. PubMed PMID: 22654785. Pubmed Central PMCID: 3360424.
18. Aminian KS, Strafella AP. Affective disorders in Parkinson's disease. *Current opinion in neurology*. 2013 Aug;26(4):339-44. PubMed PMID: 23757262.
19. Najafi MR, Chitsaz A, Askarian Z, Najafi MA. Quality of sleep in patients with Parkinson's disease. *International journal of preventive medicine*. 2013 May;4(Suppl 2):S229-33. PubMed PMID: 23776729. Pubmed Central PMCID: 3678223.
20. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases.

Journal of neurology, neurosurgery, and psychiatry. 1992 Mar;55(3):181-4. PubMed PMID: 1564476. Pubmed Central PMCID: 1014720.

21. Ministério da Saúde I. Portaria SAS/MS nº 228 - Protocolo Clínico e Diretrizes Terapêuticas: Doença de Parkinson. 2010.
22. Corti O, Lesage S, Brice A. What genetics tells us about the causes and mechanisms of Parkinson's disease. *Physiological reviews*. 2011 Oct;91(4):1161-218. PubMed PMID: 22013209.
23. Winklhofer KF, Haass C. Mitochondrial dysfunction in Parkinson's disease. *Biochimica et biophysica acta*. 2010 Jan;1802(1):29-44. PubMed PMID: 19733240.
24. Betarbet R, Sherer TB, Greenamyre JT. Ubiquitin-proteasome system and Parkinson's diseases. *Experimental neurology*. 2005 Feb;191 Suppl 1:S17-27. PubMed PMID: 15629758.
25. Berg D, Siefker C, Becker G. Echogenicity of the substantia nigra in Parkinson's disease and its relation to clinical findings. *Journal of neurology*. 2001 Aug;248(8):684-9. PubMed PMID: 11569897.
26. Bernheimer H, Birkmayer W, Hornykiewicz O, Jellinger K, Seitelberger F. Brain dopamine and the syndromes of Parkinson and Huntington. Clinical, morphological and neurochemical correlations. *Journal of the neurological sciences*. 1973 Dec;20(4):415-55. PubMed PMID: 4272516.
27. Nussbaum RL, Polymeropoulos MH. Genetics of Parkinson's disease. *Human molecular genetics*. 1997;6(10):1687-91. PubMed PMID: 9300660.
28. Halliday GM, Holton JL, Revesz T, Dickson DW. Neuropathology underlying clinical variability in patients with synucleinopathies. *Acta neuropathologica*. 2011 Aug;122(2):187-204. PubMed PMID: 21720849.
29. Spillantini MG, Schmidt ML, Lee VM, Trojanowski JQ, Jakes R, Goedert M. Alpha-synuclein in Lewy bodies. *Nature*. 1997 Aug 28;388(6645):839-40. PubMed PMID: 9278044.
30. Goedert M, Spillantini MG, Del Tredici K, Braak H. 100 years of Lewy pathology. *Nature reviews Neurology*. 2013 Jan;9(1):13-24. PubMed PMID: 23183883.

31. Olanow CW, Perl DP, DeMartino GN, McNaught KS. Lewy-body formation is an aggresome-related process: a hypothesis. *Lancet neurology*. 2004 Aug;3(8):496-503. PubMed PMID: 15261611.
32. Takahashi H, Ohama E, Suzuki S, Horikawa Y, Ishikawa A, Morita T, et al. Familial juvenile parkinsonism: clinical and pathologic study in a family. *Neurology*. 1994 Mar;44(3 Pt 1):437-41. PubMed PMID: 8145912.
33. Mori H, Kondo T, Yokochi M, Matsumine H, Nakagawa-Hattori Y, Miyake T, et al. Pathologic and biochemical studies of juvenile parkinsonism linked to chromosome 6q. *Neurology*. 1998 Sep;51(3):890-2. PubMed PMID: 9748052.
34. Tanaka M, Kim YM, Lee G, Junn E, Iwatsubo T, Mouradian MM. Aggresomes formed by alpha-synuclein and synphilin-1 are cytoprotective. *The Journal of biological chemistry*. 2004 Feb 6;279(6):4625-31. PubMed PMID: 14627698.
35. Zhou W, Freed CR. DJ-1 up-regulates glutathione synthesis during oxidative stress and inhibits A53T alpha-synuclein toxicity. *The Journal of biological chemistry*. 2005 Dec 30;280(52):43150-8. PubMed PMID: 16227205.
36. Feany MB, Bender WW. A *Drosophila* model of Parkinson's disease. *Nature*. 2000 Mar 23;404(6776):394-8. PubMed PMID: 10746727.
37. Periquet M, Fulga T, Myllykangas L, Schlossmacher MG, Feany MB. Aggregated alpha-synuclein mediates dopaminergic neurotoxicity in vivo. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2007 Mar 21;27(12):3338-46. PubMed PMID: 17376994.
38. St Martin JL, Klucken J, Outeiro TF, Nguyen P, Keller-McGandy C, Cantuti-Castelvetri I, et al. Dopaminergic neuron loss and up-regulation of chaperone protein mRNA induced by targeted over-expression of alpha-synuclein in mouse substantia nigra. *Journal of neurochemistry*. 2007 Mar;100(6):1449-57. PubMed PMID: 17241127.
39. Thomas B, Beal MF. Molecular insights into Parkinson's disease. *F1000 medicine reports*. 2011;3:7. PubMed PMID: 21655332. Pubmed Central PMCID: 3096887.
40. Samii A, Nutt JG, Ransom BR. Parkinson's disease. *Lancet*. 2004 May 29;363(9423):1783-93. PubMed PMID: 15172778.
41. Jankovic J, Poewe W. Therapies in Parkinson's disease. *Current opinion in*

neurology. 2012 Aug;25(4):433-47. PubMed PMID: 22691758.

42. Morrish PK, Rakshi JS, Bailey DL, Sawle GV, Brooks DJ. Measuring the rate of progression and estimating the preclinical period of Parkinson's disease with [18F]dopa PET. *Journal of neurology, neurosurgery, and psychiatry*. 1998 Mar;64(3):314-9. PubMed PMID: 9527140. Pubmed Central PMCID: 2170010.
43. Volkmann J. Deep brain stimulation for the treatment of Parkinson's disease. *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*. 2004 Jan-Feb;21(1):6-17. PubMed PMID: 15097290.
44. Browner N, Pagan F. Deep Brain Stimulation National Parkinson Foundation [Abril de 2013]. Available from: <http://www.parkinson.org/Parkinson-s-Disease/Treatment/Surgical-Treatment-Options/Deep-Brain-Stimulation>.
45. Duvoisin RC, Eldridge R, Williams A, Nutt J, Calne D. Twin study of Parkinson disease. *Neurology*. 1981 Jan;31(1):77-80. PubMed PMID: 7192828.
46. Langston JW, Ballard P, Tetrud JW, Irwin I. Chronic Parkinsonism in humans due to a product of meperidine-analog synthesis. *Science*. 1983 Feb 25;219(4587):979-80. PubMed PMID: 6823561.
47. Ward CD, Duvoisin RC, Ince SE, Nutt JD, Eldridge R, Calne DB. Parkinson's disease in 65 pairs of twins and in a set of quadruplets. *Neurology*. 1983 Jul;33(7):815-24. PubMed PMID: 6683366.
48. Calne S, Schoenberg B, Martin W, Uitti RJ, Spencer P, Calne DB. Familial Parkinson's disease: possible role of environmental factors. *The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques*. 1987 Aug;14(3):303-5. PubMed PMID: 3664372.
49. Tanner CM, Ottman R, Goldman SM, Ellenberg J, Chan P, Mayeux R, et al. Parkinson disease in twins: an etiologic study. *JAMA : the journal of the American Medical Association*. 1999 Jan 27;281(4):341-6. PubMed PMID: 9929087.
50. Wirdefeldt K, Gatz M, Schalling M, Pedersen NL. No evidence for heritability of Parkinson disease in Swedish twins. *Neurology*. 2004 Jul 27;63(2):305-11. PubMed PMID: 15277625.
51. Gowers W. Diseases of the nerves and spinal cord. *A Manual Of Diseases Of*

The Nervous System. I. 3rd ed. Philadelphia: Blakiston's Son & Co.; 1900.

52. Spellman GG. Report of familial cases of parkinsonism. Evidence of a dominant trait in a patient's family. JAMA : the journal of the American Medical Association. 1962 Feb 3;179:372-4. PubMed PMID: 13915636.
53. Tune LE, Folstein M, Rabins P, Jayaram G, McHugh P. Familial manic-depressive illness and familial Parkinson's disease: a case report. The Johns Hopkins medical journal. 1982 Aug;151(2):65-70. PubMed PMID: 7098246.
54. Thacker EL, Ascherio A. Familial aggregation of Parkinson's disease: a meta-analysis. Movement disorders : official journal of the Movement Disorder Society. 2008 Jun 15;23(8):1174-83. PubMed PMID: 18442112.
55. Farrer MJ. Genetics of Parkinson disease: paradigm shifts and future prospects. Nature reviews Genetics. 2006 Apr;7(4):306-18. PubMed PMID: 16543934.
56. Burbulla LF, Kruger R. Converging environmental and genetic pathways in the pathogenesis of Parkinson's disease. Journal of the neurological sciences. 2011 Jul 15;306(1-2):1-8. PubMed PMID: 21513949.
57. Langston JW, Ballard PA, Jr. Parkinson's disease in a chemist working with 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine. The New England journal of medicine. 1983 Aug 4;309(5):310. PubMed PMID: 6602944.
58. Casals J, Elizan TS, Yahr MD. Postencephalitic parkinsonism--a review. Journal of neural transmission. 1998;105(6-7):645-76. PubMed PMID: 9826109.
59. Elbaz A, Clavel J, Rathouz PJ, Moisan F, Galanaud JP, Delemotte B, et al. Professional exposure to pesticides and Parkinson disease. Annals of neurology. 2009 Oct;66(4):494-504. PubMed PMID: 19847896.
60. Gorell JM, Johnson CC, Rybicki BA, Peterson EL, Richardson RJ. The risk of Parkinson's disease with exposure to pesticides, farming, well water, and rural living. Neurology. 1998 May;50(5):1346-50. PubMed PMID: 9595985.
61. Ritz B, Ascherio A, Checkoway H, Marder KS, Nelson LM, Rocca WA, et al. Pooled analysis of tobacco use and risk of Parkinson disease. Archives of neurology. 2007 Jul;64(7):990-7. PubMed PMID: 17620489.
62. Allam MF, Campbell MJ, Del Castillo AS, Fernandez-Crehuet Navajas R. Parkinson's disease protects against smoking? Behavioural neurology.

2004;15(3-4):65-71. PubMed PMID: 15706049.

63. Allam MF, Campbell MJ, Hofman A, Del Castillo AS, Fernandez-Crehuet Navajas R. Smoking and Parkinson's disease: systematic review of prospective studies. *Movement disorders : official journal of the Movement Disorder Society*. 2004 Jun;19(6):614-21. PubMed PMID: 15197698.
64. Tanner CM, Goldman SM, Aston DA, Ottman R, Ellenberg J, Mayeux R, et al. Smoking and Parkinson's disease in twins. *Neurology*. 2002 Feb 26;58(4):581-8. PubMed PMID: 11865136.
65. Ascherio A, Weisskopf MG, O'Reilly EJ, McCullough ML, Calle EE, Rodriguez C, et al. Coffee consumption, gender, and Parkinson's disease mortality in the cancer prevention study II cohort: the modifying effects of estrogen. *American journal of epidemiology*. 2004 Nov 15;160(10):977-84. PubMed PMID: 15522854.
66. Qi H, Li S. Dose-response meta-analysis on coffee, tea and caffeine consumption with risk of Parkinson's disease. *Geriatrics & gerontology international*. 2013 Jul 23. PubMed PMID: 23879665.
67. Burn DJ, Mark MH, Playford ED, Maraganore DM, Zimmerman TR, Jr., Duvoisin RC, et al. Parkinson's disease in twins studied with 18F-dopa and positron emission tomography. *Neurology*. 1992 Oct;42(10):1894-900. PubMed PMID: 1407569.
68. Foltynie T, Sawcer S, Brayne C, Barker RA. The genetic basis of Parkinson's disease. *Journal of neurology, neurosurgery, and psychiatry*. 2002 Oct;73(4):363-70. PubMed PMID: 12235301. Pubmed Central PMCID: 1738059.
69. Piccini P, Burn DJ, Ceravolo R, Maraganore D, Brooks DJ. The role of inheritance in sporadic Parkinson's disease: evidence from a longitudinal study of dopaminergic function in twins. *Annals of neurology*. 1999 May;45(5):577-82. PubMed PMID: 10319879.
70. Hamza TH, Payami H. The heritability of risk and age at onset of Parkinson's disease after accounting for known genetic risk factors. *Journal of human genetics*. 2010 Apr;55(4):241-3. PubMed PMID: 20203693. Pubmed Central PMCID: 2947819.
71. Moilanen JS, Autere JM, Myllyla VV, Majamaa K. Complex segregation analysis of Parkinson's disease in the Finnish population. *Human genetics*. 2001

Mar;108(3):184-9. PubMed PMID: 11354627.

72. Wirdefeldt K, Gatz M, Reynolds CA, Prescott CA, Pedersen NL. Heritability of Parkinson disease in Swedish twins: a longitudinal study. *Neurobiology of aging*. 2011 Oct;32(10):1923 e1-8. PubMed PMID: 21482443.
73. International Parkinson Disease Genomics C, Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin UM, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*. 2011 Feb 19;377(9766):641-9. PubMed PMID: 21292315.
74. Barbeau A, Pourcher E. New data on the genetics of Parkinson's disease. *The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques*. 1982 Feb;9(1):53-60. PubMed PMID: 7093827.
75. Lazzarini AM, Myers RH, Zimmerman TR, Jr., Mark MH, Golbe LI, Sage JI, et al. A clinical genetic study of Parkinson's disease: evidence for dominant transmission. *Neurology*. 1994 Mar;44(3 Pt 1):499-506. PubMed PMID: 8145922.
76. Bonifati V, Fabrizio E, Vanacore N, De Mari M, Meo G. Familial Parkinson's disease: a clinical genetic analysis. *The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques*. 1995 Nov;22(4):272-9. PubMed PMID: 8599769.
77. Payami H, Bernard S, Larsen K, Kaye J, Nutt J. Genetic anticipation in Parkinson's disease. *Neurology*. 1995 Jan;45(1):135-8. PubMed PMID: 7824103.
78. Markopoulou K, Wszolek ZK, Pfeiffer RF. A Greek-American kindred with autosomal dominant, levodopa-responsive parkinsonism and anticipation. *Annals of neurology*. 1995 Sep;38(3):373-8. PubMed PMID: 7668822.
79. Plante-Bordeneuve V, Taussig D, Thomas F, Ziegler M, Said G. A clinical and genetic study of familial cases of Parkinson's disease. *Journal of the neurological sciences*. 1995 Nov;133(1-2):164-72. PubMed PMID: 8583221.
80. Marder K, Levy G, Louis ED, Mejia-Santana H, Cote L, Andrews H, et al. Familial aggregation of early- and late-onset Parkinson's disease. *Annals of neurology*. 2003 Oct;54(4):507-13. PubMed PMID: 14520664.

81. Zarepari S, Taylor TD, Harris EL, Payami H. Segregation analysis of Parkinson disease. *American journal of medical genetics*. 1998 Dec 4;80(4):410-7. PubMed PMID: 9856574.
82. Maher NE, Currie LJ, Lazzarini AM, Wilk JB, Taylor CA, Saint-Hilaire MH, et al. Segregation analysis of Parkinson disease revealing evidence for a major causative gene. *American journal of medical genetics*. 2002 May 1;109(3):191-7. PubMed PMID: 11977177.
83. McDonnell SK, Schaid DJ, Elbaz A, Strain KJ, Bower JH, Ahlskog JE, et al. Complex segregation analysis of Parkinson's disease: The Mayo Clinic Family Study. *Annals of neurology*. 2006 May;59(5):788-95. PubMed PMID: 16634030.
84. Polymeropoulos MH, Higgins JJ, Golbe LI, Johnson WG, Ide SE, Di Iorio G, et al. Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science*. 1996 Nov 15;274(5290):1197-9. PubMed PMID: 8895469.
85. Gasser T, Muller-Myhsok B, Wszolek ZK, Durr A, Vaughan JR, Bonifati V, et al. Genetic complexity and Parkinson's disease. *Science*. 1997 Jul 18;277(5324):388-9; author reply 9. PubMed PMID: 9518367.
86. Pankratz N, Nichols WC, Uniacke SK, Halter C, Rudolph A, Shults C, et al. Genome screen to identify susceptibility genes for Parkinson disease in a sample without parkin mutations. *American journal of human genetics*. 2002 Jul;71(1):124-35. PubMed PMID: 12058349. Pubmed Central PMCID: 384969.
87. Pankratz N, Nichols WC, Uniacke SK, Halter C, Rudolph A, Shults C, et al. Significant linkage of Parkinson disease to chromosome 2q36-37. *American journal of human genetics*. 2003 Apr;72(4):1053-7. PubMed PMID: 12638082. Pubmed Central PMCID: 1180337.
88. Pankratz N, Nichols WC, Uniacke SK, Halter C, Murrell J, Rudolph A, et al. Genome-wide linkage analysis and evidence of gene-by-gene interactions in a sample of 362 multiplex Parkinson disease families. *Human molecular genetics*. 2003 Oct 15;12(20):2599-608. PubMed PMID: 12925570.
89. Kurth JH, Kurth MC, Poduslo SE, Schwankhaus JD. Association of a monoamine oxidase B allele with Parkinson's disease. *Annals of neurology*. 1993 Apr;33(4):368-72. PubMed PMID: 8489207.
90. Parboosingh JS, Rousseau M, Rogan F, Amit Z, Chertkow H, Johnson WG, et al. Absence of mutations in superoxide dismutase and catalase genes in patients with Parkinson's disease. *Archives of neurology*. 1995

Dec;52(12):1160-3. PubMed PMID: 7492289.

91. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science*. 1997 Jun 27;276(5321):2045-7. PubMed PMID: 9197268.
92. Wu RM, Cheng CW, Chen KH, Lu SL, Shan DE, Ho YF, et al. The COMT L allele modifies the association between MAOB polymorphism and PD in Taiwanese. *Neurology*. 2001 Feb 13;56(3):375-82. PubMed PMID: 11171904.
93. Mattila KM, Rinne JO, Lehtimäki T, Roytta M, Ahonen JP, Hurme M. Association of an interleukin 1B gene polymorphism (-511) with Parkinson's disease in Finnish patients. *Journal of medical genetics*. 2002 Jun;39(6):400-2. PubMed PMID: 12070246. Pubmed Central PMCID: 1735144.
94. Levecque C, Elbaz A, Clavel J, Richard F, Vidal JS, Amouyel P, et al. Association between Parkinson's disease and polymorphisms in the nNOS and iNOS genes in a community-based case-control study. *Human molecular genetics*. 2003 Jan 1;12(1):79-86. PubMed PMID: 12490535.
95. Chan DK, Lam MK, Wong R, Hung WT, Wilcken DE. Strong association between N-acetyltransferase 2 genotype and PD in Hong Kong Chinese. *Neurology*. 2003 Mar 25;60(6):1002-5. PubMed PMID: 12654968.
96. Marx FP, Holzmann C, Strauss KM, Li L, Eberhardt O, Gerhardt E, et al. Identification and functional characterization of a novel R621C mutation in the synphilin-1 gene in Parkinson's disease. *Human molecular genetics*. 2003 Jun 1;12(11):1223-31. PubMed PMID: 12761037.
97. Maraganore DM, de Andrade M, Lesnick TG, Strain KJ, Farrer MJ, Rocca WA, et al. High-resolution whole-genome association study of Parkinson disease. *American journal of human genetics*. 2005 Nov;77(5):685-93. PubMed PMID: 16252231. Pubmed Central PMCID: 1271381.
98. Liu X, Cheng R, Verbitsky M, Kisselev S, Browne A, Mejia-Sanatana H, et al. Genome-wide association study identifies candidate genes for Parkinson's disease in an Ashkenazi Jewish population. *BMC medical genetics*. 2011;12:104. PubMed PMID: 21812969. Pubmed Central PMCID: 3166909.
99. Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS genetics*. 2011 Jun;7(6):e1002141. PubMed PMID: 21738487. Pubmed Central PMCID:

3121750.

100. Simon-Sanchez J, van Hilten JJ, van de Warrenburg B, Post B, Berendse HW, Arepalli S, et al. Genome-wide association study confirms extant PD risk loci among the Dutch. *European journal of human genetics : EJHG*. 2011 Jun;19(6):655-61. PubMed PMID: 21248740. Pubmed Central PMCID: 3110043.
101. Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature genetics*. 2009 Dec;41(12):1308-12. PubMed PMID: 19915575. Pubmed Central PMCID: 2787725.
102. Saad M, Lesage S, Saint-Pierre A, Corvol JC, Zelenika D, Lambert JC, et al. Genome-wide association study confirms BST1 and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. *Human molecular genetics*. 2011 Feb 1;20(3):615-27. PubMed PMID: 21084426.
103. Edwards TL, Scott WK, Almonte C, Burt A, Powell EH, Beecham GW, et al. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. *Annals of human genetics*. 2010 Mar;74(2):97-109. PubMed PMID: 20070850. Pubmed Central PMCID: 2853717.
104. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nature genetics*. 2009 Dec;41(12):1303-7. PubMed PMID: 19915576.
105. Fung HC, Scholz S, Matarin M, Simon-Sanchez J, Hernandez D, Britton A, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet neurology*. 2006 Nov;5(11):911-6. PubMed PMID: 17052657.
106. Latourelle JC, Pankratz N, Dumitriu A, Wilk JB, Goldwurm S, Pezzoli G, et al. Genomewide association study for onset age in Parkinson disease. *BMC medical genetics*. 2009;10:98. PubMed PMID: 19772629. Pubmed Central PMCID: 2758866.
107. Pankratz N, Wilk JB, Latourelle JC, DeStefano AL, Halter C, Pugh EW, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Human genetics*. 2009 Jan;124(6):593-605. PubMed PMID: 18985386. Pubmed Central PMCID: 2627511.

108. Metzker ML. Sequencing technologies - the next generation. *Nature reviews Genetics*. 2010 Jan;11(1):31-46. PubMed PMID: 19997069.
109. Morey M, Fernandez-Marmiesse A, Castineiras D, Fraga JM, Couce ML, Cocho JA. A glimpse into past, present, and future DNA sequencing. *Molecular genetics and metabolism*. 2013 Sep-Oct;110(1-2):3-24. PubMed PMID: 23742747.
110. Zimprich A, Benet-Pages A, Struhal W, Graf E, Eck SH, Offman MN, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *American journal of human genetics*. 2011 Jul 15;89(1):168-75. PubMed PMID: 21763483. Pubmed Central PMCID: 3135812.
111. Vilarino-Guell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, Lincoln SJ, et al. VPS35 mutations in Parkinson disease. *American journal of human genetics*. 2011 Jul 15;89(1):162-7. PubMed PMID: 21763482. Pubmed Central PMCID: 3135796.
112. Lubbe S, Morris HR. Recent advances in Parkinson's disease genetics. *Journal of neurology*. 2013 Jun 25. PubMed PMID: 23798000.
113. Fujioka S, Wszolek ZK. Update on genetics of parkinsonism. *Neuro-degenerative diseases*. 2012;10(1-4):257-60. PubMed PMID: 22261420. Pubmed Central PMCID: 3363351.
114. Matsumine H, Saito M, Shimoda-Matsubayashi S, Tanaka H, Ishikawa A, Nakagawa-Hattori Y, et al. Localization of a gene for an autosomal recessive form of juvenile Parkinsonism to chromosome 6q25.2-27. *American journal of human genetics*. 1997 Mar;60(3):588-96. PubMed PMID: 9042918. Pubmed Central PMCID: 1712507.
115. Gasser T, Muller-Myhsok B, Wszolek ZK, Oehlmann R, Calne DB, Bonifati V, et al. A susceptibility locus for Parkinson's disease maps to chromosome 2p13. *Nature genetics*. 1998 Mar;18(3):262-5. PubMed PMID: 9500549.
116. Farrer M, Gwinn-Hardy K, Muentner M, DeVrieze FW, Crook R, Perez-Tur J, et al. A chromosome 4p haplotype segregating with Parkinson's disease and postural tremor. *Human molecular genetics*. 1999 Jan;8(1):81-5. PubMed PMID: 9887334.
117. Leroy E, Boyer R, Auburger G, Leube B, Ulm G, Mezey E, et al. The ubiquitin pathway in Parkinson's disease. *Nature*. 1998 Oct 1;395(6701):451-2. PubMed PMID: 9774100.

118. Valente EM, Bentivoglio AR, Dixon PH, Ferraris A, Ialongo T, Frontali M, et al. Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35-p36. *American journal of human genetics*. 2001 Apr;68(4):895-900. PubMed PMID: 11254447. Pubmed Central PMCID: 1275643.
119. van Duijn CM, Dekker MC, Bonifati V, Galjaard RJ, Houwing-Duistermaat JJ, Snijders PJ, et al. Park7, a novel locus for autosomal recessive early-onset parkinsonism, on chromosome 1p36. *American journal of human genetics*. 2001 Sep;69(3):629-34. PubMed PMID: 11462174. Pubmed Central PMCID: 1235491.
120. Funayama M, Hasegawa K, Kowa H, Saito M, Tsuji S, Obata F. A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. *Annals of neurology*. 2002 Mar;51(3):296-301. PubMed PMID: 11891824.
121. Hampshire DJ, Roberts E, Crow Y, Bond J, Mubaidin A, Wriekat AL, et al. Kufor-Rakeb syndrome, pallido-pyramidal degeneration with supranuclear upgaze paresis and dementia, maps to 1p36. *Journal of medical genetics*. 2001 Oct;38(10):680-2. PubMed PMID: 11584046. Pubmed Central PMCID: 1734748.
122. Hicks AA, Petursson H, Jonsson T, Stefansson H, Johannsdottir HS, Sainz J, et al. A susceptibility gene for late-onset idiopathic Parkinson's disease. *Annals of neurology*. 2002 Nov;52(5):549-55. PubMed PMID: 12402251.
123. Strauss KM, Martins LM, Plun-Favreau H, Marx FP, Kautzmann S, Berg D, et al. Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson's disease. *Human molecular genetics*. 2005 Aug 1;14(15):2099-111. PubMed PMID: 15961413.
124. Gao X, Martin ER, Liu Y, Mayhew G, Vance JM, Scott WK. Genome-wide linkage screen in familial Parkinson disease identifies loci on chromosomes 3 and 18. *American journal of human genetics*. 2009 Apr;84(4):499-504. PubMed PMID: 19327735. Pubmed Central PMCID: 2667976.
125. Shojaei S, Sina F, Banihosseini SS, Kazemi MH, Kalhor R, Shahidi GA, et al. Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. *American journal of human genetics*. 2008 Jun;82(6):1375-84. PubMed PMID: 18513678. Pubmed Central PMCID: 2427312.
126. Chartier-Harlin MC, Dachsel JC, Vilarino-Guell C, Lincoln SJ, Leprete F, Hulihan MM, et al. Translation initiator EIF4G1 mutations in familial Parkinson

disease. American journal of human genetics. 2011 Sep 9;89(3):398-406. PubMed PMID: 21907011. Pubmed Central PMCID: 3169825.

127. Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide BM, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. PLoS genetics. 2012;8(3):e1002548. PubMed PMID: 22438815. Pubmed Central PMCID: 3305333.
128. Pittman A, Hardy J. Genetic analysis in neurology: the next 10 years. JAMA neurology. 2013 Jun;70(6):696-702. PubMed PMID: 23571731.
129. Klein C, Westenberger A. Genetics of Parkinson's disease. Cold Spring Harbor perspectives in medicine. 2012 Jan;2(1):a008888. PubMed PMID: 22315721. Pubmed Central PMCID: 3253033.
130. Ishikawa A, Tsuji S. Clinical analysis of 17 patients in 12 Japanese families with autosomal-recessive type juvenile parkinsonism. Neurology. 1996 Jul;47(1):160-6. PubMed PMID: 8710071.
131. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature. 1998 Apr 9;392(6676):605-8. PubMed PMID: 9560156.
132. Hattori N, Kitada T, Matsumine H, Asakawa S, Yamamura Y, Yoshino H, et al. Molecular genetic analysis of a novel Parkin gene in Japanese families with autosomal recessive juvenile parkinsonism: evidence for variable homozygous deletions in the Parkin gene in affected individuals. Annals of neurology. 1998 Dec;44(6):935-41. PubMed PMID: 9851438.
133. Abbas N, Lucking CB, Ricard S, Durr A, Bonifati V, De Michele G, et al. A wide variety of mutations in the parkin gene are responsible for autosomal recessive parkinsonism in Europe. French Parkinson's Disease Genetics Study Group and the European Consortium on Genetic Susceptibility in Parkinson's Disease. Human molecular genetics. 1999 Apr;8(4):567-74. PubMed PMID: 10072423.
134. West AB, Maraganore D, Crook J, Lesnick T, Lockhart PJ, Wilkes KM, et al. Functional association of the parkin gene promoter with idiopathic Parkinson's disease. Human molecular genetics. 2002 Oct 15;11(22):2787-92. PubMed PMID: 12374768.
135. Scott WK, Nance MA, Watts RL, Hubble JP, Koller WC, Lyons K, et al. Complete genomic screen in Parkinson disease: evidence for multiple genes.

JAMA : the journal of the American Medical Association. 2001 Nov 14;286(18):2239-44. PubMed PMID: 11710888.

136. Zhang Y, Gao J, Chung KK, Huang H, Dawson VL, Dawson TM. Parkin functions as an E2-dependent ubiquitin- protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1. Proceedings of the National Academy of Sciences of the United States of America. 2000 Nov 21;97(24):13354-9. PubMed PMID: 11078524. Pubmed Central PMCID: 27228.
137. Shimura H, Hattori N, Kubo S, Mizuno Y, Asakawa S, Minoshima S, et al. Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. Nature genetics. 2000 Jul;25(3):302-5. PubMed PMID: 10888878.
138. Perez FA, Palmiter RD. Parkin-deficient mice are not a robust model of parkinsonism. Proceedings of the National Academy of Sciences of the United States of America. 2005 Feb 8;102(6):2174-9. PubMed PMID: 15684050. Pubmed Central PMCID: 548311.
139. Ebersole TA, Chen Q, Justice MJ, Artzt K. The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. Nature genetics. 1996 Mar;12(3):260-5. PubMed PMID: 8589716.
140. Bockbrader K, Feng Y. Essential function, sophisticated regulation and pathological impact of the selective RNA-binding protein QKI in CNS myelin development. Future neurology. 2008 Nov;3(6):655-68. PubMed PMID: 19727426. Pubmed Central PMCID: 2630251.
141. Lockhart PJ, O'Farrell CA, Farrer MJ. It's a double knock-out! The quaking mouse is a spontaneous deletion of parkin and parkin co-regulated gene (PACRG). Movement disorders : official journal of the Movement Disorder Society. 2004 Jan;19(1):101-4. PubMed PMID: 14743368.
142. Lorenzetti D, Antalffy B, Vogel H, Noveroske J, Armstrong D, Justice M. The neurological mutant quaking(viable) is Parkin deficient. Mammalian genome : official journal of the International Mammalian Genome Society. 2004 Mar;15(3):210-7. PubMed PMID: 15014970.
143. Aberg K, Saetre P, Lindholm E, Ekholm B, Pettersson U, Adolfsson R, et al. Human QKI, a new candidate gene for schizophrenia involved in myelination. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2006 Jan 5;141B(1):84-90. PubMed PMID: 16342280.

144. Chenard CA, Richard S. New implications for the QUAKE RNA binding protein in human disease. *Journal of neuroscience research*. 2008 Feb 1;86(2):233-42. PubMed PMID: 17787018.
145. Lauriat TL, Shiue L, Haroutunian V, Verbitsky M, Ares M, Jr., Ospina L, et al. Developmental expression profile of quaking, a candidate gene for schizophrenia, and its target genes in human prefrontal cortex and hippocampus shows regional specificity. *Journal of neuroscience research*. 2008 Mar;86(4):785-96. PubMed PMID: 17918747.
146. Coordinators NR. Database resources of the National Center for Biotechnology Information. *Nucleic acids research*. 2013 Jan;41(Database issue):D8-D20. PubMed PMID: 23193264. Pubmed Central PMCID: 3531099.
147. National Center for Biothecnology Information (NCBI). Available from: <http://www.ncbi.nlm.nih.gov>.
148. Ensembl. Available from: <http://www.ensembl.org>.
149. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, et al. The Ensembl genome database project. *Nucleic acids research*. 2002 Jan 1;30(1):38-41. PubMed PMID: 11752248. Pubmed Central PMCID: 99161.
150. Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, et al. Ensembl 2013. *Nucleic acids research*. 2013 Jan;41(Database issue):D48-55. PubMed PMID: 23203987. Pubmed Central PMCID: 3531136.
151. International HapMap C. The International HapMap Project. *Nature*. 2003 Dec 18;426(6968):789-96. PubMed PMID: 14685227.
152. Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site. *Genome research*. 2005 Nov;15(11):1592-3. PubMed PMID: 16251469. Pubmed Central PMCID: 1310647.
153. Backx L, Fryns JP, Marcelis C, Devriendt K, Vermeesch J, Van Esch H. Haploinsufficiency of the gene Quaking (QKI) is associated with the 6q terminal deletion syndrome. *American journal of medical genetics Part A*. 2010 Feb;152A(2):319-26. PubMed PMID: 20082458.
154. West AB, Lockhart PJ, O'Farrell C, Farrer MJ. Identification of a novel gene linked to parkin via a bi-directional promoter. *Journal of molecular biology*. 2003 Feb 7;326(1):11-9. PubMed PMID: 12547187.

155. Taylor JM, Song YJ, Huang Y, Farrer MJ, Delatycki MB, Halliday GM, et al. Parkin Co-Regulated Gene (PACRG) is regulated by the ubiquitin-proteasomal system and is present in the pathological features of Parkinsonian diseases. *Neurobiology of disease*. 2007 Aug;27(2):238-47. PubMed PMID: 17590346.
156. Taylor JM, Wu RM, Farrer MJ, Delatycki MB, Lockhart PJ. Analysis of Parkin Co-Regulated Gene in a Taiwanese-ethnic Chinese cohort with early-onset Parkinson's disease. *Parkinsonism & related disorders*. 2009 Jul;15(6):417-21. PubMed PMID: 19196541.
157. Imai Y, Soda M, Murakami T, Shoji M, Abe K, Takahashi R. A product of the human gene adjacent to parkin is a component of Lewy bodies and suppresses Pael receptor-induced cell death. *The Journal of biological chemistry*. 2003 Dec 19;278(51):51901-10. PubMed PMID: 14532270.
158. The Human Gene Compendium (GeneCards): Weizmann Institute of Science. Available from: <http://www.genecards.org>.
159. Kawajiri S, Saiki S, Sato S, Sato F, Hatano T, Eguchi H, et al. PINK1 is recruited to mitochondria with parkin and associates with LC3 in mitophagy. *FEBS letters*. 2010 Mar 19;584(6):1073-9. PubMed PMID: 20153330.
160. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009 Oct 8;461(7265):747-53. PubMed PMID: 19812666. Pubmed Central PMCID: 2831613.
161. Maher B. Personal genomes: The case of the missing heritability. *Nature*. 2008 Nov 6;456(7218):18-21. PubMed PMID: 18987709.
162. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *The New England journal of medicine*. 2010 Apr 1;362(13):1181-91. PubMed PMID: 20220177.
163. Hoischen A, van Bon BW, Gilissen C, Arts P, van Lier B, Steehouwer M, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nature genetics*. 2010 Jun;42(6):483-5. PubMed PMID: 20436468.
164. Lalonde E, Albrecht S, Ha KC, Jacob K, Bolduc N, Polychronakos C, et al. Unexpected allelic heterogeneity and spectrum of mutations in Fowler syndrome revealed by next-generation exome sequencing. *Human mutation*. 2010 Aug;31(8):918-23. PubMed PMID: 20518025.

165. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics*. 2010 Jan;42(1):30-5. PubMed PMID: 19915526. Pubmed Central PMCID: 2847889.
166. Walsh T, Shahin H, Elkan-Miller T, Lee MK, Thornton AM, Roeb W, et al. Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein GPSM2 as the cause of nonsyndromic hearing loss DFNB82. *American journal of human genetics*. 2010 Jul 9;87(1):90-4. PubMed PMID: 20602914. Pubmed Central PMCID: 2896776.
167. Byun M, Abhyankar A, Lelarge V, Plancoulaine S, Palanduz A, Telhan L, et al. Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *The Journal of experimental medicine*. 2010 Oct 25;207(11):2307-12. PubMed PMID: 20876309. Pubmed Central PMCID: 2964585.
168. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Current opinion in genetics & development*. 2009 Jun;19(3):212-9. PubMed PMID: 19481926. Pubmed Central PMCID: 2914559.
169. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *Journal of genetics and genomics = Yi chuan xue bao*. 2011 Mar 20;38(3):95-109. PubMed PMID: 21477781. Pubmed Central PMCID: 3076108.
170. Novelli G, Predazzi IM, Mango R, Romeo F, Mehta JL. Role of genomics in cardiovascular medicine. *World journal of cardiology*. 2010 Dec 26;2(12):428-36. PubMed PMID: 21191544. Pubmed Central PMCID: 3011138.
171. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature reviews Genetics*. 2010 Jun;11(6):415-25. PubMed PMID: 20479773.
172. Mardis ER. Next-generation DNA sequencing methods. *Annual review of genomics and human genetics*. 2008;9:387-402. PubMed PMID: 18576944.
173. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 2011 Jul 21;475(7356):348-52. PubMed PMID: 21776081.
174. Pavlopoulos GA, Oulas A, Iacucci E, Sifrim A, Moreau Y, Schneider R, et al. Unraveling genomic variation from next generation sequencing data. *BioData*

- mining. 2013;6(1):13. PubMed PMID: 23885890. Pubmed Central PMCID: 3726446.
175. Chan EY. Next-generation sequencing methods: impact of sequencing accuracy on SNP discovery. *Methods in molecular biology*. 2009;578:95-111. PubMed PMID: 19768588.
 176. Lupton MK, Proitsi P, Danillidou M, Tsolaki M, Hamilton G, Wroe R, et al. Deep sequencing of the Nicastrin gene in pooled DNA, the identification of genetic variants that affect risk of Alzheimer's disease. *PloS one*. 2011;6(2):e17298. PubMed PMID: 21364883. Pubmed Central PMCID: 3045431.
 177. Edwards MD, Gifford DK. High-resolution genetic mapping with pooled sequencing. *BMC bioinformatics*. 2012;13 Suppl 6:S8. PubMed PMID: 22537047. Pubmed Central PMCID: 3358661.
 178. Chen Q, Sun F. A unified approach for allele frequency estimation, SNP detection and association studies based on pooled sequencing data using EM algorithms. *BMC genomics*. 2013;14 Suppl 1:S1. PubMed PMID: 23369070. Pubmed Central PMCID: 3549804.
 179. Smith AM, Heisler LE, St Onge RP, Farias-Hesson E, Wallace IM, Bodeau J, et al. Highly-multiplexed barcode sequencing: an efficient method for parallel analysis of pooled samples. *Nucleic acids research*. 2010 Jul;38(13):e142. PubMed PMID: 20460461. Pubmed Central PMCID: 2910071.
 180. John SW, Weitzner G, Rozen R, Sriver CR. A rapid procedure for extracting genomic DNA from leukocytes. *Nucleic acids research*. 1991 Jan 25;19(2):408. PubMed PMID: 2014181. Pubmed Central PMCID: 333617.
 181. Genome Reference Consortium Human Build 37 patch release 10 (GRCh37.p10) [Internet]. Genome Reference Consortium. 2012. Available from: http://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.22/.
 182. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--new capabilities and interfaces. *Nucleic acids research*. 2012 Aug;40(15):e115. PubMed PMID: 22730293. Pubmed Central PMCID: 3424584.
 183. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology*. 2000;132:365-86. PubMed PMID: 10547847.

184. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics*. 2007 May 15;23(10):1289-91. PubMed PMID: 17379693.
185. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome research*. 2002 Jun;12(6):996-1006. PubMed PMID: 12045153. Pubmed Central PMCID: 186604.
186. Incorporated LT. Ion Xpress Plus gDNA and amplicon library preparation. User's guide. Version number 4471989, revision B. Life Technologies Incorporated. 2011.
187. Casals F, Idaghdour Y, Hussin J, Awadalla P. Next-generation sequencing approaches for genetic mapping of complex diseases. *Journal of neuroimmunology*. 2012 Jul 15;248(1-2):10-22. PubMed PMID: 22285396.
188. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome research*. 1998 Mar;8(3):186-94. PubMed PMID: 9521922.
189. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome research*. 1998 Mar;8(3):175-85. PubMed PMID: 9521921.
190. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nature biotechnology*. 2011 Jan;29(1):24-6. PubMed PMID: 21221095. Pubmed Central PMCID: 3346182.
191. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*. 2013 Mar;14(2):178-92. PubMed PMID: 22517427. Pubmed Central PMCID: 3603213.
192. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9. PubMed PMID: 19505943. Pubmed Central PMCID: 2723002.
193. Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS computational biology*. 2013 Apr;9(4):e1003031. PubMed PMID: 23592973. Pubmed Central PMCID: 3623719.

194. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, et al. Database resources of the National Center for Biotechnology Information. *Nucleic acids research*. 2007 Jan;35(Database issue):D5-12. PubMed PMID: 17170002. Pubmed Central PMCID: 1781113.
195. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010 Oct 28;467(7319):1061-73. PubMed PMID: 20981092. Pubmed Central PMCID: 3042601.
196. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012 Nov 1;491(7422):56-65. PubMed PMID: 23128226. Pubmed Central PMCID: 3498066.
197. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nature methods*. 2010 Apr;7(4):248-9. PubMed PMID: 20354512. Pubmed Central PMCID: 2855889.
198. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics* / editorial board, Jonathan L Haines [et al]. 2013 Jan;Chapter 7:Unit7 20. PubMed PMID: 23315928.
199. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*. 2003 Jul 1;31(13):3812-4. PubMed PMID: 12824425. Pubmed Central PMCID: 168916.
200. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols*. 2009;4(7):1073-81. PubMed PMID: 19561590.
201. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research*. 2009 May;37(9):e67. PubMed PMID: 19339519. Pubmed Central PMCID: 2685110.
202. Wang M, Marin A. Characterization and prediction of alternative splice sites. *Gene*. 2006 Feb 1;366(2):219-27. PubMed PMID: 16226402.
203. Chang TH, Huang HY, Hsu JB, Weng SL, Horng JT, Huang HD. An enhanced computational platform for investigating the roles of regulatory RNA and for

- identifying functional RNA motifs. *BMC bioinformatics*. 2013;14 Suppl 2:S4. PubMed PMID: 23369107. Pubmed Central PMCID: 3549854.
204. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*. 2004;32(5):1792-7. PubMed PMID: 15034147. Pubmed Central PMCID: 390337.
 205. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*. 2004 Aug 19;5:113. PubMed PMID: 15318951. Pubmed Central PMCID: 517706.
 206. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*. 2011;7:539. PubMed PMID: 21988835. Pubmed Central PMCID: 3261699.
 207. Stenson PD, Ball EV, Mort M, Phillips AD, Shaw K, Cooper DN. The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]*. 2012 Sep;Chapter 1:Unit1 13. PubMed PMID: 22948725.
 208. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, et al. Human Gene Mutation Database (HGMD): 2003 update. *Human mutation*. 2003 Jun;21(6):577-81. PubMed PMID: 12754702.
 209. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Human mutation*. 2011 May;32(5):557-63. PubMed PMID: 21520333.
 210. Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biology of the cell / under the auspices of the European Cell Biology Organization*. 2009 May;101(5):251-62. PubMed PMID: 19275763.
 211. Trempe JF, Sauve V, Grenier K, Seirafi M, Tang MY, Menade M, et al. Structure of parkin reveals mechanisms for ubiquitin ligase activation. *Science*. 2013 Jun 21;340(6139):1451-5. PubMed PMID: 23661642.
 212. West A, Periquet M, Lincoln S, Lucking CB, Nicholl D, Bonifati V, et al. Complex relationship between Parkin mutations and Parkinson disease. *American journal of medical genetics*. 2002 Jul 8;114(5):584-91. PubMed PMID: 12116199.

213. Sutherland G, Mellick G, Sue C, Chan DK, Rowe D, Silburn P, et al. A functional polymorphism in the parkin gene promoter affects the age of onset of Parkinson's disease. *Neuroscience letters*. 2007 Mar 6;414(2):170-3. PubMed PMID: 17280783.
214. Ross OA, Haugarvoll K, Stone JT, Heckman MG, White LR, Aasly JO, et al. Lack of evidence for association of Parkin promoter polymorphism (PRKN-258) with increased risk of Parkinson's disease. *Parkinsonism & related disorders*. 2007 Oct;13(7):386-8. PubMed PMID: 17400506.
215. Taylor JM, Wu RM, Lin CH, Delatycki MB, Lockhart PJ. Lack of evidence for association of a parkin promoter polymorphism with early-onset Parkinson's disease in a Chinese population. *Parkinsonism & related disorders*. 2009 Feb;15(2):149-52. PubMed PMID: 18387843.
216. Chang XL, Mao XY, Li HH, Zhang JH, Li NN, Burgunder JM, et al. Functional parkin promoter polymorphism in Parkinson's disease: new data and meta-analysis. *Journal of the neurological sciences*. 2011 Mar 15;302(1-2):68-71. PubMed PMID: 21176923.
217. Foroud T, Uniacke SK, Liu L, Pankratz N, Rudolph A, Halter C, et al. Heterozygosity for a mutation in the parkin gene leads to later onset Parkinson disease. *Neurology*. 2003 Mar 11;60(5):796-801. PubMed PMID: 12629236.
218. Hedrich K, Marder K, Harris J, Kann M, Lynch T, Meija-Santana H, et al. Evaluation of 50 probands with early-onset Parkinson's disease for Parkin mutations. *Neurology*. 2002 Apr 23;58(8):1239-46. PubMed PMID: 11971093.
219. Trempe JF, Fon EA. Structure and Function of Parkin, PINK1, and DJ-1, the Three Musketeers of Neuroprotection. *Frontiers in neurology*. 2013;4:38. PubMed PMID: 23626584. Pubmed Central PMCID: 3630392.
220. Rose JM, Novoselov SS, Robinson PA, Cheetham ME. Molecular chaperone-mediated rescue of mitophagy by a Parkin RING1 domain mutant. *Human molecular genetics*. 2011 Jan 1;20(1):16-27. PubMed PMID: 20889486. Pubmed Central PMCID: 3000674.
221. Lucking CB, Chesneau V, Lohmann E, Verpillat P, Dulac C, Bonnet AM, et al. Coding polymorphisms in the parkin gene and susceptibility to Parkinson disease. *Archives of neurology*. 2003 Sep;60(9):1253-6. PubMed PMID: 12975291.
222. Gaweda-Walerych K, Safranow K, Jasinska-Myga B, Bialecka M, Klodowska-

Duda G, Rudzinska M, et al. PARK2 variability in Polish Parkinson's disease patients--interaction with mitochondrial haplogroups. *Parkinsonism & related disorders*. 2012 Jun;18(5):520-4. PubMed PMID: 22361577. Pubmed Central PMCID: 3358581.

223. Clark LN, Afridi S, Karlins E, Wang Y, Mejia-Santana H, Harris J, et al. Case-control study of the parkin gene in early-onset Parkinson disease. *Archives of neurology*. 2006 Apr;63(4):548-52. PubMed PMID: 16606767.
224. Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature structural & molecular biology*. 2009 Feb;16(2):107-13. PubMed PMID: 19190664.

APÊNDICES

APÊNDICE A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

(Com autorização para armazenamento)

"Análise por sequenciamento comparativo direto do gene *QKI* em pacientes com Doença de Parkinson de Início Precoce"

A pontifícia Universidade Católica do Paraná (PUCPR) estará realizando um estudo científico sobre Doença de Parkinson. Você pode estar sendo convidado a participar do estudo por ser portador de Doença de Parkinson ou para compor o grupo controle, formado por indivíduos saudáveis. No estudo, pretende-se avaliar o papel de um gene como possível candidato a participar das causas da Doença de Parkinson de Início Precoce. Se bem sucedido o projeto poderá facilitar o entendimento da Doença de Parkinson e auxiliar em tratamentos melhores futuros. Este estudo será coordenado pelo Dr. Marcelo Távora Mira, professor adjunto e pesquisador em atividade na Escola de Saúde e Biociências da PUCPR e o farmacêutico Wilian Corrêa de Macedo mestrando do Programa de Pós-Graduação em Genética da Universidade Federal do Paraná.

1) Procedimentos

Se você concordar em participar deste estudo, será submetido a exame médico e à coleta de uma amostra de sangue. Serão coletados 5.0mL (um tubo pequeno) de sangue de seu antebraço, de maneira idêntica àquela realizada nos laboratórios de análises clínicas para exame de sangue. O sangue será utilizado para extração do ácido nucléico (DNA). O DNA extraído será armazenado até o momento da genotipagem. O DNA será submetido a um teste de pesquisa chamado PCR (Reação em Cadeia da Polimerase) para análise do gene em estudo. Informações contidas no seu prontuário médico também poderão ser lidas pelos pesquisadores e utilizadas no estudo.

2) Permissão para armazenamento

Está sendo solicitada a sua permissão para estocagem da sua amostra de DNA na PUCPR. Se assinar esse termo de consentimento, você está autorizando estocagem de longo prazo de sua amostra de DNA. Sua amostra de DNA será mantida indefinidamente. Isto significa que sua amostra não será destruída após um determinado período de tempo, mas sim será estocada pelo tempo que ela durar. Amostras estocadas serão exclusivamente para fins de pesquisa. Amostras estocadas poderão ser utilizadas para estudos adicionais de susceptibilidade à Doença de Parkinson. Nenhuma outra doença será investigada. O uso da

sua amostra estocada terá como condição uma nova avaliação e aprovação do projeto de pesquisa pelo Comitê de Ética pertinente.

3) Local de estudo

Estes procedimentos serão realizados parte no Núcleo de Aconselhamento e Laboratório de Genética – GENETIKA, parte no Núcleo de Investigação Molecular Avançada da PUCPR, Campus Curitiba, e parte no Laboratório de Genômica e Biologia Molecular da PUCPR Campus São José dos Pinhais. Estas parcerias entre a PUCPR e outras instituições de pesquisa podem aumentar ainda mais a possibilidade de sucesso e a importância dos resultados. Você autoriza a utilização do material coletado em projetos de pesquisa em parceria com outras instituições nacionais, desde que mantidos o desenho, objetivos e metodologia do projeto original, e que as alterações propostas sejam avaliadas e aprovadas pelo CEP.

4) Riscos/Desconfortos

Os riscos para a saúde de participação neste estudo são muito pequenos e limitados ao procedimento de coleta de sangue. Durante a coleta de sangue, você poderá sentir um desconforto temporário devido a introdução da agulha. A coleta de sangue poderá resultar em uma pequena lesão que quase sempre cura-se sozinha. Em raros casos, pode ocorrer infecção localizada.

5) Tratamento e Compensação por Danos

Se você desenvolver infecção localizada devido ao procedimento de coleta de sangue, o tratamento será providenciado pela PUCPR. O custo deste tratamento será totalmente coberto pelo projeto.

6) Alternativas

Você tem total liberdade para decidir ou não pela participação neste estudo. Caso você decida não participar, ou cessar sua participação no estudo a qualquer momento, esta decisão não irá interferir de nenhuma forma em qualquer procedimento médico para diagnóstico ou tratamento de Doença de Parkinson ou qualquer outra doença que você possa necessitar no futuro. Da mesma forma, sua decisão não irá refletir no acesso a procedimentos médicos necessários a algum familiar ou contato.

7) Custos para os participantes

No caso de você decidir participar do estudo, você não terá nenhum custo. Custos com testes laboratoriais e análise de suas amostras para propósito de pesquisa serão cobertos pelo estudo.

8) Benefícios

Em longo prazo, os procedimentos médicos e laboratoriais aos quais você será submetido poderão facilitar a detecção de resistência ao tratamento de doença de Parkinson, tornando possível evitar tratamentos inadequados. Além disso, espera-se que conhecimentos científicos adicionais sejam alcançados, com consequente melhoria do tratamento de pessoas afetadas pela Doença de Parkinson.

9) Reembolso

Você não será reembolsado por participar deste estudo.

10) Confidencialidade dos dados

Amostras de DNA e células SERÃO UTILIZADAS APENAS PARA PESQUISA DE SUSCEPTIBILIDADE À DOENÇA DE PARKINSON. Nenhuma outra doença será estudada. Se, devido a situações imprevistas, os investigadores descobrirem alguma informação relacionada a susceptibilidade a outras doenças que não a Doença de Parkinson, você será aconselhado a procurar seu médico para um exame completo. Você não terá acesso a nenhuma informação genética específica que nos produzirmos. Todos os resultados obtidos no estudo, após análise do conjunto completo dos dados, serão publicados em artigos científicos. Como estes dados publicados serão usados por outros investigadores é desconhecido. Importante reafirmar, o alvo de nossos estudos é a indentificação do fator de risco genético à Doença de Parkinson.

11) Confidenciabilidade dos dados

A participação em projetos de pesquisa deste tipo pode resultar em perda da privacidade. Além disso, a descoberta de fatores de risco genético para a Doença de Parkinson pode expor susceptibilidade de certos grupos de pessoas, possivelmente levando a outros ou a certas empresas a considerar estes grupos diferentes de uma forma negativa. Entretanto, procedimentos serão adotados pelos responsáveis por este estudo no intuito de proteger a confidencialidade das informações que você fornecer e as informações produzidas pelo projeto. Nenhuma informação genética individual será tornada pública. As informações serão codificadas e mantidas num local reservado o tempo todo. Somente os pesquisadores envolvidos neste estudo terão acesso às informações. Após o término deste estudo, as informações serão transcritas para os arquivos de computador, mantidos em local restrito com acesso permitido apenas aos mesmos pesquisadores. Os dados deste estudo poderão ser discutidos com pesquisadores de outras instituições, mas nenhuma identificação será fornecida.

CONSENTIMENTO

Você receberá uma cópia deste consentimento para mantê-lo consigo. Nos próximos dias, se tiver qualquer dúvida sobre a sua participação neste estudo, você poderá utilizar os seguintes meios de contato com o pesquisador responsável.

Dr. Marcelo Távora Mira
Telefone: (41) 3271-2030
E-mail: m.mira@pucpr.br

Farm. Wilian Corrêa de Macedo
Telefone: (41) 9814-5623
E-mail: wilian.macedo@hotmail.com

Em caso de reclamação ou qualquer tipo de denúncia sobre este estudo devo ligar para o CEP PUCPR (41) 3271-2292 ou mandar um email para nep@pucpr.br.

A PARTICIPAÇÃO EM PESQUISA É VOLUNTÁRIA

Você tem o direito de não concordar em participar ou mesmo de se retirar do estudo em qualquer momento que queira, sem riscos para o seu tratamento médico, ou de familiares. Se você desejar e concordar em participar, deve assinar na linha abaixo.

Se você desejar participar do estudo, permitirá também que seu endereço e telefone sejam anotados em uma folha separada, para facilitar contato quando necessário. Como já foi esclarecida anteriormente, toda informação pessoal será mantida em sigilo.

Assinatura do voluntário

Nome completo e N° do prontuário

Assinatura do entrevistador

Nome completo do entrevistador

Assinatura testemunha 1

Assinatura testemunha 2

Data

APÊNDICE B - FERRAMENTAS ONLINE UTILIZADAS

TABELA B-1. Ferramentas online utilizadas neste estudo.

Ferramenta	URL	Referências
Análise de conservação		
Clustal Omega	http://www.clustal.org/omega/	Sievers, 2011
MUSCLE (<i>Multiple Sequence Comparison by Log-Expectation</i>)	http://www.drive5.com/muscle/	Edgar, 2004; Edgar, 2004
Base de dados		
HGMD (<i>Human Gene Mutation Database</i>)	http://www.hgmd.org/	Stenson, 2003; Stenson, 2012
LOVD (<i>Leiden Open Variation Database</i>)	http://www.lovd.nl/3.0/home	Fokkema, 2011
PDgene database	http://www.pdgene.org/	Lill, 2012
Desenho de primers		
Primer3	http://bioinfo.ut.ee/primer3-0.4.0/	Koressaar, 2007; Untergasser, 2012
Genome browser		
1000 Genomes	http://www.1000genomes.org/	Genomes Project Consortium, 2010; Genomes Project Consortium, 2012
HapMap (<i>International HapMap Project</i>)	http://hapmap.ncbi.nlm.nih.gov/	International HapMap Consortium, 2003; Thorisson, 2005
IGV (<i>Integrative Genome Viewer</i>)	http://www.broadinstitute.org/igv/	Robinson, 2011; Thorvaldsdottir, 2013
Múltiplas funções		
Ensembl	http://www.ensembl.org/index.html	Hubbard, 2002; Flicek, 2013
NCBI (<i>National Center for Biotechnology Information</i>)	http://www.ncbi.nlm.nih.gov/	NCBI Resource Coordinators, 2013
UCSC Genome Browser	http://genome.ucsc.edu/	Kent, 2002
Predição de impacto funcional		
ASSP (<i>Alternative Splice Site Predictor</i>)	http://wangcomputing.com/assp/index.html	Wang, 2006
HSF (<i>Human Splice Finder</i>)	http://www.umd.be/HSF/	Desmet, 2009
PolyPhen 2 (<i>Polymorphism Phenotyping 2</i>)	http://genetics.bwh.harvard.edu/pph2/index.shtml	Adzhubei, 2010; Adzhubei, 2013
RegRNA	http://regrna2.mbc.nctu.edu.tw/	Chang, 2013
SIFT (<i>Sorting Intolerant From Tolerant</i>)	http://sift.jcvi.org/	Ng, 2003; Kumar, 2009

APÊNDICE C - PRIMERS DESENHADOS PARA OS GENES *PACRG*, *PARK2* e *QKI***TABELA C-1.** Informações dos primers desenhados para o gene *PACRG*.

Amplicon	Gene/Exon	Sequencia do iniciador	Posição (GRCh37.p10)	Extensão do amplicon gerado	Extensão do Exon
1	<i>PACRG/1</i>	CCCCATCTAGCACTGTTTT CGGTGCCAACTGTCAAT	chr6:163147843-163148605	763 pb	148 pb
2	<i>PACRG/2</i>	GGAGGCTTACCTTTGGAA CGGTCCAGAAACCATAAGT	chr6:163149018-163149675	658 pb	232 pb
3	<i>PACRG/3</i>	CTTTCCCAGGTGTTTGTATT GCAAAGCAACTCTGTCATT	chr6:163235007-163235533	527 pb	135 pb
4	<i>PACRG/4</i>	GGTCTAAAGGTCCATGATTTT GTCTGGCCTCCTAAGTCAA	chr6:163482974-163483495	522 pb	172 pb
5	<i>PACRG/5</i>	CAATGTGTCATCCAAAGCTA GCTGTACCTGAGGAGTAATAAA	chr6:163510135-163510652	518p pb	150 pb
6	<i>PACRG/6</i>	CGAACACCATGTCAATCTT CTGTCACAGGTAAAGCTGAA	chr6:163733612-163734222	611 pb	117 pb
7	<i>PACRG/7</i>	CCGACTTGTCATCTTTCTTG GGTTTAACTCGGACTCCTG	chr6:163735591-163736303	713 pb	665 pb
8	<i>PACRG/7</i>	CATTAAGTACGTGGTCCCAAC CCATACGGTGGAGAGACAAG	chr6:163735974-163736763	790 pb	

TABELA C-2.Informações dos primersdesenhados para o gene *PARK2*.

Amplicon	Gene/Exon	Sequencia do iniciador	Posição (GRCh37.p10)	Extensão do amplicon gerado	Extensão do Exon
1	<i>PARK2</i> /1	GTAAGATTGCTGGGCCTGAA TTCGAACGCACACACTGG	chr6:163148495-163148962	468 pb	110 pb
2	<i>PARK2</i> /2	GTTTGCAGGTCACTGACGAA TCCAATCTTTCCTGCTTGCT	chr6:162864144-162864734	591 pb	164 pb
3	<i>PARK2</i> /3	GTTACATGTCACTTTTGCTTCCCTTC GTACTCCACCTACAGTGATGTCTCCTTG	chr6:162683400-162683913	514 pb	241 pb
4	<i>PARK2</i> /4	TTCAAAATGTTCTGAAGAC TTTCTTTTCAAAGACGGGTG	chr6:162622058-162622402	345 pb	122 pb
5	<i>PARK2</i> /5	TGGAAACATGTCTTAAGGAGTACA TGGCAAACATTATTAGATGGAAGA	chr6:162475003-162475277	275 pb	84 pb
6	<i>PARK2</i> /6	ACTTTGGCACAAGGTCATCC GGGCAGACGCATCAAAATAA	chr6:162394167-162394623	457 pb	116 pb
7	<i>PARK2</i> /7	GCCTCCAATTTTTAAGATGT CAATTCCTTCATTCCCCAGA	chr6:162206620-162207116	497 pb	137 pb
8	<i>PARK2</i> /8	GTCTGTGTCCTCCCAAAGG TGGTTTTCTTCCCCATTTCA	chr6:161990258-161990563	306 pb	62 pb
9	<i>PARK2</i> /9	TGCCAGTTGAATTTGGTAGA GCAACCAAAGCAGAAAATCT	chr6:161969657-161970256	600 pb	150 pb
10	<i>PARK2</i> /10	CCCATTCCCATGGCTGTC TGAACAATCCCATCAAAACA	chr6:161807689-161808081	393 pb	84 pb
11	<i>PARK2</i> /11	CCATCCGCCTAGTAGCTGTC GGAATCCCTGATGGGTATGA	chr6:161781047-161781407	361 pb	118 pb
12	<i>PARK2</i> /12	GTTTGGGAATGCGTGTTTTC GCTCTGCTGTCTTGTGTGGA	chr6:161770868-161771312	445 pb	2792 pb

TABELA C-3. Informações dos *primers* desenhados para o gene *QKI*.

Amplicon	Gene/Exon	Sequencia do iniciador	Coordenada hg19	Extensão do amplicon gerado	Extensão do Exon
1	<i>QKI</i> /1	GCCAAGTGAGTGAGGAGGAG TCCTGGCCTCCCAGTCC	chr6:163835696-163836500	805 pb	693 pb
2	<i>QKI</i> /2	TTTTCCTCTGGACTAGCAAGAAT TGAAGCAGAACAAAATTGAATAAAA	chr6:163876246-163876495	250 pb	143 pb
3	<i>QKI</i> /3	TGTAAGTTCCAGCAAGTGAATAATG AAAGACTTAGTCCATTTTCAAGGA	chr6:163899759-163899957	199 pb	117 pb
4	<i>QKI</i> /4	GGTTAGTATTATGAAAGATGTGCTCTC GCCCAAATCTAAGGTCTGGAA	chr6:163955941-163956190	250 pb	144 pb
5	<i>QKI</i> /5	TTTGTGGACAGTCTCAAATGCT CAACGCAGACAGCAGTAACC	chr6:163982965-163983145	181b pb	88 pb
6	<i>QKI</i> /6	CTCCCTTCTCATTTTTCCGTA CCCATGGGGAGAAGAACTTA	chr6:163984380-163984772	393 pb	300 pb
7	<i>QKI</i> /7	CCCCCTTGTGAATGTTTTCT TCAGGTGTGGTGTGCAAAT	chr6:163987702-163987901	200 pb	75 pb
8	<i>QKI</i> /8	TGGTGTCAATCAGTTTGTGAGA TCTCTCATGGCACACTTTCTTC	chr6:163991566-163992368	803 pb	7903 pb

APÊNDICE D - TEMPERATURA EXPERIMENTAL DE HIBRIDAÇÃO DOS PARES DE *PRIMERS*

TABELA D-1. Temperatura experimental de hibridação dos pares de *primers* desenhados neste estudo.

<i>PARK2</i>	Temperatura	<i>PACRG</i>	Temperatura	<i>QKI</i>	Temperatura
Exon 1	55°C	Exon 1	54°C	Exon 1	55°C
Exon 2	55°C	Exon 2	57°C	Exon 2	55°C
Exon 3	56°C	Exon 3	57°C	Exon 3	55°C
Exon 4	55°C	Exon 4	58°C	Exon 4	57°C
Exon 5	55°C	Exon 5	59°C	Exon 5	57°C
Exon 6	55°C	Exon 6	56°C	Exon 6	57°C
Exon 7	55°C	Exon 7	59°C	Exon 7	55°C
Exon 8	55°C	Exon 7	60°C	Exon 8	55°C
Exon 9	55°C				
Exon 10	55°C				
Exon 11	55°C				
Exon 12	55°C				

ANEXOS

ANEXO A - EXTRAÇÃO DE DNA PELA TÉCNICA DE *SALTING OUT*

1. Coletar de 5 a 10 mL de sangue (material biológico) com EDTA (Anti-coagulante);
2. Centrifugar por 10 minutos;
3. Pipetar camada intermediária *Buffy-coat* e passar para um tubo Falcon de 15mL;
4. Adiciona-se 6mL de TKM1 (Tabela A);
5. Acrescentar 100 μ L de igepal (Nonidet);
6. Agitar no Vortex para misturar bem o detergente;
7. Completar até 14mL com TKM1, agitar no vortex por 2 minutos;
8. Centrifugar por 10 minutos a 3.000rpm;
9. Desprezar o sobrenadante deixando somente o pellet;
10. Completar até 6mL com TKM1, suspender o pellet e agitar no vortex;
11. Completar até 14mL com TKM1 e agitar novamente no vortex;
12. Centrifugar por 10 minutos a 3.000rpm;
13. Desprezar o sobrenadante deixando somente o pellet;
14. Completar até 14mL com TKM1, ressuspender o pellet e agitar no vortex;
15. Centrifugar por 10 minutos a 3.000rpm;
16. Acrescentar 800 μ L de TKM2 (Tabela B);
17. Ressuspender o pellet com pipeta de Pasteur;
18. Passar para o eppendorf de 1,5mL;
19. Adicionar 50 μ L de SDS 10% a 55°C e agitar no vortex;
20. Incubar em banho-maria a 55°C por 24 horas (*overnight*);
21. Retirar do banho e adicionar 300 μ L de NaCl saturado 6M (4M – 5M) e agitar no vortex;
22. Centrifugar por 15 minutos a 12.000rpm;
23. Transferir o sobrenadante para um tubo de ensaio, no sobrenadante está o DNA e no precipitado, as proteínas;
24. Adicionar 2 volumes de etanol 100% a –20°C (2mL);
25. Tampar com parafilme e inverter várias vezes;
26. Pescar o DNA com a ponteira;
27. Passar para um eppendorf e lavar o DNA em etanol 70% a 4°C (1mL);
28. Centrifugar 10 minutos a 13.400 rpm. O DNA é insolúvel em álcool e os 30% de água possibilitam a solubilização do restante dos sais presentes no tubo;
29. Retirar o etanol 70% e deixar secar em temperatura ambiente (Estufa);
30. Aplicar 80 μ L de TE (Tabela C);
31. Incubar em banho-maria de 1 a 2 dias até o DNA entrar em solução com o TE;
32. Colocar na geladeira por 1 semana;
33. Passar para o freezer.

TABELA A. Solução de TKM1.

Reagentes	Concentração estoque	Volume
Tris-HCl pH 7,6	1M	5mL
KCl	1M	5mL
MgCl ₂	1M	5mL
EDTA	0,1M	10mL
ddH ₂ O	q.s.p.	500mL

TABELA B. Solução de TKM2.

Reagentes	Concentração estoque	Volume
Tris-HCl pH 7,6	1M	5mL
KCl	1M	5mL
MgCl ₂	1M	5mL
EDTA	0,1M	10mL
ddH ₂ O	q.s.p.	500mL

TABELA C. Solução de TE.

Reagentes	Concentração estoque	Volume
Tris-HCl pH 8,0	1M	0,5mL
EDTA	0,1M	0,5mL
ddH ₂ O	q.s.p.	50mL